

```

XX PR 03-AUG-1998; 98US-0127815.
XX XX
XX PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX XX
XX PI Fasano A;
XX XX
XX DR WPI: 2000-205565/18.
XX XX
PT New peptide antagonist of zonulin useful as antiinflammatory agent for
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis
XX XX
PS Claim 1: Page 42; 69pp; English.
XX XX
CC This present sequence is that of a peptide antagonist of zonulin
CC (2), one of 25 such peptides (see Y79105-29) of the invention,
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not
CC physiologically modulate the opening of mammalian tight junctions
CC (TJ). The peptide antagonists are based on a common motif of ZOT
CC and human zonulins, which is believed to be critical for receptor
CC binding. They can be prepared by chemical synthesis or by use of
CC recombinant DNA techniques. The peptide antagonists are used as an
CC antiinflammatory agents in the treatment of gastrointestinal
CC inflammation, where they bind to the ZOT receptor in the intestine
CC and yet does not physiologically modulate the opening of TJ in the
CC intestine. Gastrointestinal inflammation conditions give rise to
CC increased intestinal permeability and the peptide is useful for
CC treating intestinal conditions that cause protein losing enteropathy
CC caused by infection, e.g. Clostridium difficile infection,
CC enterocolitis, shigellosis, viral gastroenteritis, parasite
CC infestation, bacterial overgrowth, whipple's disease, diseases with
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,
CC collagenous colitis, inflammatory bowel disease, diseases marked by
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,
CC sarcolosis lymphoma, mesenteric tuberculosis, and after surgical
CC correction of congenital heart disease with Fontan's operation,
CC mucosal diseases without ulceration, e.g. Menetrier's disease,
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,
CC e.g. systemic lupus erythematosus or food allergies, primarily to
CC milk.
XX XX
SO Sequence 8 AA:
QY 1 GRLLVQPG 8
DB 1 grllvqpg 8
Query Match 87.8%; Score 36; DB 21; Length 8;
Best Local Similarity 87.5%; Pred. No. 3.2e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

FT FT /note="complementarity determining region 1"
FT FT 50..66
FT FT /label= CDR2
FT FT /note="complementarity determining region 2"
FT FT 99..107
FT FT /label= CDR3
FT FT /note="complementarity determining region 3"
XX XX
PN WO9720932-A1.
XX XX
PD 12-JUN-1997.
XX XX
PF 09-DEC-1996; 96WO-GB03043.
XX XX
PR 11-OCT-1996; 96GB-0021295.
PR 07-DEC-1995; 95GB-0025004.
PR 23-MAY-1996; 96GB-0010824.
XX XX
PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
XX XX
PI Allen DJ, McCafferty JG, Osbourn JK;
XX XX
DR WPI: 1997-319779/29.
DR N-PSDB; T72128.
XX XX
PT Specific binding members for human carcinoembryonic antigen - bind
PT to the A3-B3 extracellular domain of hCEA and are substantially
PT non-cross-reactive with human liver cells; used for diagnosing
PT cancer
XX XX
PS Claim 7: Fig 1a; 128pp; English.
XX XX
CC This polypeptide sequence comprises the heavy chain variable
CC region (VH) of human carcinoembryonic antigen (hCEA)-specific
CC antibody CEA3. VH (T72126-32) and VL (T72133-35) gene sequences
CC were obtained for anti-hCEA antibodies CEA1-CEA7 (see W19876-85).
CC A claimed specific binding member (A) comprises an hCEA specific
CC antibody antigen binding domain that has a dissociation constant
CC for hCEA of less than 1 x 10-8 M, is non-cross-reactive with human
CC liver cells, and preferentially binds to the A3-B3 extracellular
CC domain of hCEA and/or to cell-associated hCEA over hCEA over
CC soluble hCEA. Preferred (A) include pairings of VH and VL
CC sequences from CEA1-7, or their CDR sequences, as well as CEA6
CC VH and VL variants. (A) is used to detect cells expressing hCEA,
CC in vivo or in vitro, especially tumour cells for diagnosing cancer,
CC e.g. adenocarcinoma of the colon, lung or breast.
XX XX
SO Sequence 118 AA:
QY 1 GRLLVQPG 8
DB 8 grllvqpg 15
Query Match 87.8%; Score 36; DB 18; Length 118;
Best Local Similarity 87.5%; Pred. No. 8.7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

RESULT 4
W19878
ID W19878 standard; Protein: 118 AA.
XX XX
XX AC W19878;
XX XX
DT 07-DEC-1997 (first entry)
XX XX
DE CEA-specific antibody CEA3 VH sequence.
XX XX
XX DE Carcinoembryonic antigen; CEA: human; antibody; scFv;
XX XX
XX KM tumour marker; lung cancer; breast cancer; colon cancer;
XX XX
XX KM adenocarcinoma; diagnosis.
XX XX
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX XX 31..35
XX FT Region /label= CDR1

```

```

RESULT 5
G20127
ID G20127 standard; Protein: 314 AA.
XX XX
XX AC G20127;
XX XX
DT 17-OCT-2000 (first entry)
XX XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 22193.
XX XX
XX XX Protein identification; signal transduction pathway; metabolic pathway;
XX XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX XX termination sequence.

```

```
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137532.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139454.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145819.
PR 27-JUL-1999; 99US-0145819.
PR 28-JUL-1999; 99US-0145819.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
```

PR 15-SEP-1999; 99US-0154018-  
PR 16-SEP-1999; 99US-0154039-  
PR 20-SEP-1999; 99US-0154779-  
PR 22-SEP-1999; 99US-0155139-  
PR 23-SEP-1999; 99US-0155486-  
PR 24-SEP-1999; 99US-0155659-  
PR 28-SEP-1999; 99US-0156458-  
PR 29-SEP-1999; 99US-0156596-  
PR 04-OCT-1999; 99US-0157117-  
PR 05-OCT-1999; 99US-0157753-  
PR 06-OCT-1999; 99US-0157865-  
PR 07-OCT-1999; 99US-0158029-  
PR 08-OCT-1999; 99US-0158232-  
PR 12-OCT-1999; 99US-0158369-  
PR 13-OCT-1999; 99US-0158293-  
PR 13-OCT-1999; 99US-0159294-  
PR 13-OCT-1999; 99US-0159295-  
PR 14-OCT-1999; 99US-0159329-  
PR 14-OCT-1999; 99US-0159330-  
PR 14-OCT-1999; 99US-0159331-  
PR 14-OCT-1999; 99US-0159637-  
PR 14-OCT-1999; 99US-0159638-  
PR 18-OCT-1999; 99US-0159584-  
PR 21-OCT-1999; 99US-0160741-  
PR 21-OCT-1999; 99US-0160767-  
PR 21-OCT-1999; 99US-0160768-  
PR 21-OCT-1999; 99US-0160770-  
PR 21-OCT-1999; 99US-0160814-  
PR 21-OCT-1999; 99US-0160815-  
PR 22-OCT-1999; 99US-0160980-  
PR 22-OCT-1999; 99US-0160981-  
PR 22-OCT-1999; 99US-0160989-  
PR 25-OCT-1999; 99US-0161404-  
PR 25-OCT-1999; 99US-0161405-  
PR 25-OCT-1999; 99US-0161406-  
PR 26-OCT-1999; 99US-0161359-  
PR 26-OCT-1999; 99US-0161360-  
PR 26-OCT-1999; 99US-0161361-  
PR 28-OCT-1999; 99US-0161920-  
PR 28-OCT-1999; 99US-0161993-  
PR 28-OCT-1999; 99US-0162142-  
PR 29-OCT-1999; 99US-0162142-  
Query Match 87.8%; Score 36; DB 21; Length 314;  
Best Local Similarity 87.5%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLLVPG 8  
      |||||  
Db 29 grllvspg 36

RESULT 6  
G20126  
ID G20126 standard; Protein: 321 AA.  
XX G20126;  
AC  
XX 17-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 22192.  
DE  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
PN EPI033405-A2.  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF

XX  
PR 25-FEB-1999; 99US-0121825-  
PR 05-MAR-1999; 99US-0123180-  
PR 09-MAR-1999; 99US-0123548-  
PR 23-MAR-1999; 99US-0125788-  
PR 25-MAR-1999; 99US-0126264-  
PR 29-MAR-1999; 99US-0126785-  
PR 01-APR-1999; 99US-0127462-  
PR 06-APR-1999; 99US-0128234-  
PR 08-APR-1999; 99US-0128714-  
PR 16-APR-1999; 99US-0129845-  
PR 19-APR-1999; 99US-0130077-  
PR 21-APR-1999; 99US-0130449-  
PR 23-APR-1999; 99US-0130510-  
PR 23-APR-1999; 99US-0130891-  
PR 28-APR-1999; 99US-0131449-  
PR 30-APR-1999; 99US-0132048-  
PR 30-APR-1999; 99US-0132407-  
PR 04-MAY-1999; 99US-0132484-  
PR 05-MAY-1999; 99US-0132485-  
PR 06-MAY-1999; 99US-0132486-  
PR 06-MAY-1999; 99US-0132487-  
PR 07-MAY-1999; 99US-0132863-  
PR 11-MAY-1999; 99US-0134256-  
PR 14-MAY-1999; 99US-0134218-  
PR 14-MAY-1999; 99US-0134221-  
PR 14-MAY-1999; 99US-0134370-  
PR 18-MAY-1999; 99US-0134768-  
PR 19-MAY-1999; 99US-0134941-  
PR 20-MAY-1999; 99US-0135124-  
PR 21-MAY-1999; 99US-0135353-  
PR 24-MAY-1999; 99US-0135629-  
PR 25-MAY-1999; 99US-0136021-  
PR 27-MAY-1999; 99US-0136392-  
PR 28-MAY-1999; 99US-0136782-  
PR 01-JUN-1999; 99US-0137222-  
PR 03-JUN-1999; 99US-0137528-  
PR 04-JUN-1999; 99US-0137502-  
PR 07-JUN-1999; 99US-0137724-  
PR 08-JUN-1999; 99US-0138094-  
PR 10-JUN-1999; 99US-0138540-  
PR 14-JUN-1999; 99US-0138847-  
PR 16-JUN-1999; 99US-0139119-  
PR 16-JUN-1999; 99US-0139452-  
PR 16-JUN-1999; 99US-0139453-  
PR 17-JUN-1999; 99US-0139492-  
PR 18-JUN-1999; 99US-0139454-  
PR 18-JUN-1999; 99US-0139455-  
PR 18-JUN-1999; 99US-0139456-  
PR 18-JUN-1999; 99US-0139457-  
PR 18-JUN-1999; 99US-0139458-  
PR 18-JUN-1999; 99US-0139459-  
PR 18-JUN-1999; 99US-0139460-  
PR 18-JUN-1999; 99US-0139461-  
PR 18-JUN-1999; 99US-0139462-  
PR 18-JUN-1999; 99US-0139463-  
PR 18-JUN-1999; 99US-0139750-  
PR 18-JUN-1999; 99US-0139763-  
PR 21-JUN-1999; 99US-0139817-  
PR 22-JUN-1999; 99US-0139899-  
PR 23-JUN-1999; 99US-0140353-  
PR 23-JUN-1999; 99US-0140354-  
PR 24-JUN-1999; 99US-0140695-  
PR 28-JUN-1999; 99US-0140823-  
PR 29-JUN-1999; 99US-0140991-  
PR 30-JUN-1999; 99US-0141287-  
PR 01-JUL-1999; 99US-0141842-  
PR 01-JUL-1999; 99US-0142154-  
PR 02-JUL-1999; 99US-0142055-  
PR 06-JUL-1999; 99US-0142390-  
PR 08-JUL-1999; 99US-0142803-  
PR 09-JUL-1999; 99US-0142920-

```

PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144633.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145089.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145226.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147058.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 13-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.

```

```

PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159283.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 25-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

```

```

Query Match      87.8%; Score 36; DB 21; Length 321;
Best Local Similarity 87.5%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 GRLLVOPG 8
Db 36 grllvsppg 43

```

```

RESULT 7
G20125
ID G20125 standard; Protein: 461 AA.
AC
XX G20125;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22191.
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 22191.
XX
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.

```



PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149724.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

—

PR	23-APR-1999	9905-0110891
PR	28-APR-1999	9905-0131469
PR	30-APR-1999	9905-0132008
PR	30-APR-1999	9905-0132407
PR	04-MAY-1999	9905-0132440
PR	05-MAY-1999	9905-0132465
PR	06-MAY-1999	9905-0132465
PR	06-MAY-1999	9905-0132465
PR	07-MAY-1999	9905-0132467
PR	11-MAY-1999	9905-0134256
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134221
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134491
PR	20-MAY-1999	9905-0135144
PR	21-MAY-1999	9905-0135353
PR	25-MAY-1999	9905-0135639
PR	27-MAY-1999	9905-0136021
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137282
PR	03-JUN-1999	9905-0137502
PR	04-JUN-1999	9905-0137524
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138069
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	17-JUN-1999	9905-0139453
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139753
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139899
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140605
PR	26-JUN-1999	9905-0140823
PR	30-JUN-1999	9905-0140991
PR	01-JUL-1999	9905-0141287
PR	01-JUL-1999	9905-0141482
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142554
PR	06-JUL-1999	9905-0143624
PR	08-JUL-1999	9905-0143290
PR	09-JUL-1999	9905-0142603
PR	12-JUL-1999	9905-0142920
PR	13-JUL-1999	9905-0143477
PR	14-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0144054
PR	15-JUL-1999	9905-0144065
PR	16-JUL-1999	9905-0144086
PR	16-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	19-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144631

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147459.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 314;  
Best Local Similarity 87.5%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVPG 8  
| | | | |  
Db 29 grlvtvpg 36

## RESULT 9

G48785  
ID G48785 standard; Protein: 321 AA.

AC G48785;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 61643.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0128845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.



PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 85.4%; Score 35; DB 21; Length 321;  
Best Local Similarity 87.5%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 GRLVOPC 8  
Db 36 grlvtppg 43  
  
RESULT 10  
G48784  
ID G48784 standard; Protein; 461 AA.  
XX G48784;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61642.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.

```

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

```

```

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 85.4%; Score 35; DB 21; Length 461;  
 Best Local Similarity 87.5%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GRLVQPG 8  
 Db 176 grllvpg 183

RESULT 11  
 Y37055  
 ID Y37055 standard; Protein: 599 AA.

AC Y37055;

DT 07-OCT-1999 (first entry)

DE Amino acid sequence of a Chlamydia trachomatis protein.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perithenitis;  
 KW nonnococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumonia; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

PN WO9928475-A2.

PD 10-JUN-1999.

PF 27-NOV-1998; 98WO-IB01939.

PR 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

PI Griffals R;

DR WPI: 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

PS Disclosure; Page 874; 1755pp; English.

CC Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of  
 CC Chlamydia trachomatis (see 201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nonnococcal urethritis, epididymitis, cervicitis, salpingitis;  
 CC perithenitis, Bartholinitis; pneumonia in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.

XX Sequence 599 AA.

Query Match 85.4%; Score 35; DB 20; Length 599;  
 Best Local Similarity 62.5%; Pred. No. 65;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
 DB 438 grlllepg 445

RESULT 12  
 ID Y26579 standard; Protein; 1122 AA.  
 Y26579

AC Y26579;  
 DT 13-SEP-1999 (first entry)  
 DE Murine telomerase reverse transcriptase (mTERT) enzyme.  
 XX  
 XX Telomerase reverse transcriptase; TERT; mouse; telomere length assay;  
 KW Immunogen; enzyme; telomerase-mediated DNA replication.  
 KM  
 XX  
 OS Mus sp.  
 PN WO9927113-A1.  
 PD 03-JUN-1999.  
 PF 25-NOV-1998; 98WO-US25211.  
 XX  
 XX 16-MAR-1998; 98US-0042460.  
 PR 26-NOV-1997; 97US-0979742.  
 XX  
 PA (GERO-) GERON CORP.  
 PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.  
 PI Allsopp R, Deinbo R, Greenberg R, Morin GB;  
 XX  
 XX WPI: 1999-347722/29.  
 DR N-PSDB; X80994.  
 DR  
 XX  
 PT Mouse telomerase reverse transcriptase (mTERT) enzyme proteins and  
 PT nucleic acids  
 PS  
 PS Claim 8; Fig 2; 135pp; English.  
 XX  
 CC The invention relates to a mouse telomerase reverse transcriptase (mTERT)  
 CC enzyme. Compositions containing mTERT can be used in telomere length  
 CC assays. Isolated mTERT is useful as an immunogen for the production of  
 CC monoclonal or polyclonal antibodies. The method is useful for assessing  
 CC the degree of purification and identification of new mTERT species, such  
 CC as an mTERT allele, homolog or isoform, or to screen for modulators  
 CC (antagonists and agonists) of telomerase-mediated DNA replication.  
 CC Antagonists and agonists of mTERT can be used to modify the activity of  
 CC other telomerase enzymes such as human TERT (hTERT). The present sequence  
 CC represents a mTERT enzyme.  
 CC  
 SO Sequence 1122 AA;

Query Match 85.4%; Score 35; DB 20; Length 1122;  
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
 DB 35 grllvpg 42

RESULT 13  
 ID Y79123 standard; Peptide; 8 AA.  
 Y79123;  
 AC Y79123;  
 XX  
 XX 05-JUN-2000 (first entry)  
 DT  
 XX

DE Peptide antagonist of zonulin.  
 XX  
 XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antileuc; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 KW  
 XX  
 OS Synthetic.  
 PN WO200007609-A1.  
 PD 17-FEB-2000.  
 PF 28-JUL-1999; 99WO-US16683.  
 XX  
 XX 03-AUG-1998; 98US-0127815.  
 PR  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 PI Fasano A;  
 XX  
 XX WPI: 2000-205565/18.  
 DR  
 XX  
 XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
 XX  
 XX  
 PS Claim 1; Page 47; 69pp; English.

This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 SO Sequence 8 AA;

Query Match 82.9%; Score 34; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
 DB 1 gglllvpg 8

RESULT 14  
 ID Y34750 standard; Protein; 538 AA.  
 Y34750  
 XX

```

XX  Y34750;
AC
XX
DT  13-SEP-1999 (first entry)
DE
XX  Amino acid sequence of a Chlamydia pneumoniae protein.
XX
XX  Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX  sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX  vaccine; neutralising epitope.
XX
XX  Chlamydia pneumoniae.
XX
XX  WO9227105-A2.
XX
XX  03-JUN-1999.
XX
XX  20-NOV-1998; 98WO-IB01890.
XX
XX  04-NOV-1998; 98US-0107078.
XX  21-NOV-1997; 97FR-0014673.
XX
XX  (GEST ) GENSET.
XX
XX  Griffiths R;
XX
XX  MPI; 1999-357842/30.
XX
XX  Genome sequence of Chlamydia pneumoniae
XX
XX  Page 737-738; Disclosure: 1912pp; English.
XX
XX  Y34584-Y35879 represent the proteins encoded by all the open reading
XX  frames in the complete genome (see X91990) of Chlamydia pneumoniae.
XX  C. pneumoniae causes respiratory disease such as pneumonia and
XX  bronchitis and is thought to be a contributing factor in heart
XX  disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX  nodosum or pharyngitis. The polypeptides encoded by the open reading
XX  frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
XX  immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX  nucleotide sequences can also be used as immunogenic compositions,
XX  especially where the vector directs the expression of a neutralising
XX  epitope of C. pneumoniae.
XX
XX  Sequence 538 AA;
XX
SQ

```

```

Query Match      82.9%; Score 34; DB 20; Length 538;
Best Local Similarity 62.5%; Pred. No. 91;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

OY  1 GRLLVOPG 8
    |||::||
DB  466 grlvlepg 473

```

```

RESULT 15
Y79105
ID  Y79105 standard; Peptide: 8 AA.
XX
XX  Y79105;
AC
XX
XX  05-JUN-2000 (first entry)
DE
XX
XX  Peptide antagonist of zonulin.
XX
XX  Zonulin; antagonist; zonula occludens toxin receptor;
XX  blood-brain barrier; antiinflammatory; cerebroprotective;
XX  neuroprotective; dermatologic; antidiarr; antiviral;
XX  antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;
XX  hypotensive; immunosuppressive; antiparasitic; vasotropic;
XX  gastrointestinal inflammation; therapy.
XX

```

```

OS  Synthetic.
XX
XX  WO200007609-A1.
XX
XX  17-FEB-2000.
XX
XX  28-JUL-1999; 99WO-US16683.
XX
XX  03-AUG-1998; 98US-0127815.
XX
XX  (UYMA-) UNIV MARYLAND BALTIMORE.
XX
XX  Fasano A;
XX
XX  MPI; 2000-205565/18.
XX
XX  New peptide antagonist of zonulin useful as antiinflammatory agent for
XX  treating cerebral ischemia, stroke, cerebral edema, gastritis,
XX  shigellosis, viral gastroenteritis, meningitis, encephalomyelitis
XX
XX  Claim 1, Page 41; 69pp; English.
XX

```

This present sequence is that of a peptide antagonist of zonulin (Z), one of 25 such peptides (see Y79105-29) of the invention, which bind to a zonula occludens toxin (ZOT) receptor, yet do not physiologically modulate the opening of mammalian tight junctions (TJ). The peptide antagonists are based on a common motif of ZOT and human zonulins, which is believed to be critical for receptor binding. They can be prepared by chemical synthesis or by use of recombinant DNA techniques. The peptide antagonists are used as an antiinflammatory agents in the treatment of gastrointestinal inflammation, where they bind to the ZOT receptor in the intestine and yet does not physiologically modulate the opening of TJ in the intestine. Gastrointestinal inflammation conditions give rise to increased intestinal permeability and the peptide is useful for treating intestinal conditions that cause protein losing enteropathy caused by infection, e.g. Clostridium difficile infection, enterocolitis, shigellosis, viral gastroenteritis, parasite infestation, bacterial overgrowth, whipple's disease, diseases with mucosal erosion or ulcerations, e.g. gastritis, gastric cancer, collagenous colitis, inflammatory bowel disease, diseases marked by lymphatic obstruction, e.g. congenital intestinal lymphangiectasia, sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical correction of congenital heart disease with Fontan's operation, CC mucosal diseases without ulceration, e.g. Menetrier's disease, CC coeliac disease, eosinophilic gastroenteritis, and immune diseases, CC e.g. systemic lupus erythematosus or food allergies, primarily to milk.

```

XX  Sequence 8 AA;
XX
SQ

```

```

Query Match      80.5%; Score 33; DB 21; Length 8;
Best Local Similarity 75.0%; Pred. No. 3.2e+05;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY  1 GRLLVOPG 8
    ||:||||
DB  1 grvcvpg 8

```

```

Search completed: June 13, 2001, 14:14:33
Job time: 374 sec

```





**This Page Blank (u:pj:u)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:40 ; Search time 130.61 Seconds  
(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-7

Sequence: 1 GRLLVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	614	2 B71551	probable s/t prote
2	35	85.4	614	2 H81703	conserved hypothet
3	34	82.9	358	2 T36415	probable iron-side
4	34	82.9	427	2 D83347	probable aminotran
5	34	82.9	588	2 T35549	hypothetical prote
6	34	82.9	619	2 D81556	conserved hypothet
7	34	82.9	619	2 A72114	s/t protein kinase
8	33	80.5	200	2 T29058	probable transcrip
9	33	80.5	255	2 G83014	probable nuclease
10	33	80.5	268	2 T17319	hypothetical prote
11	33	80.5	526	2 F83166	hypothetical prote
12	33	80.5	569	2 A43317	germ cell-less pro
13	33	80.5	659	1 B44212	structural protein
14	33	80.5	660	1 VHWMH2	structural protein
15	33	80.5	1230	2 T30517	complement C3-Q2 -
16	32	78.0	108	2 T70697	omega protein - hu
17	32	78.0	354	2 E82850	fibrinial adhesin p
18	32	78.0	391	2 C70972	probable trehalose
19	32	78.0	485	2 PC4427	Nck, Ash and phosph
20	32	78.0	919	2 T37062	probable transcrip
21	32	78.0	1050	2 T43482	hypothetical prote
22	31	75.6	132	1 G69256	conserved hypothet
23	31	75.6	203	2 C64724	protein catE - Esc
24	31	75.6	208	2 S27657	hypothetical prote
25	31	75.6	253	2 B82748	ubiquinol cytochro
26	31	75.6	286	2 T35657	formamidopyrimidin
27	31	75.6	396	2 D26956	translational elonga
28	31	75.6	403	2 E71300	probable protein-g
29	31	75.6	448	2 A57435	3,4-dihydroxypheny

30	31	75.6	539	2 I46470	estrogen dependent
31	31	75.6	657	2 E70529	hypothetical prote
32	31	75.6	662	2 D83367	hypothetical prote
33	31	75.6	680	2 T27078	hypothetical prote
34	31	75.6	918	2- I58178	glutamate receptor
35	31	75.6	1400	2 B70963	hypothetical prote
36	31	75.6	1440	2 T44872	probable integral
37	31	75.6	1875	2 A36429	integrin beta-4 ch
38	31	75.6	4342	2 H83343	probable non-ribos
39	30	73.2	85	2 I64783	hypothetical prote
40	30	73.2	104	2 A58158	hypothetical prote
41	30	73.2	104	2 S50341	hypothetical prote
42	30	73.2	169	2 C75595	hypothetical prote
43	30	73.2	171	1 D69986	conserved hypothet
44	30	73.2	175	1 D69522	conserved hypothet
45	30	73.2	267	2 B82694	copper homeostasis

## ALIGNMENTS

RESULT 1  
B71551  
probable s/t protein kinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: B71551  
R:Stephens, R.S.; Kalman, S.; Lamell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: B71551  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-614 <ARN>  
A:Cross-references: GB:AE001288; GB:AE001273; NID:g3328545; PIDN:AAC67736.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: pkn1

Query Match 85.4%; Score 35; DB 2; Length 614;  
Best Local Similarity 62.5%; Pred. No. 21;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVQPG 8  
DB 453 GRLLVQPG 460

RESULT 2  
H81703  
conserved hypothetical protein TC0422 [imported] - Chlamydia muridarum (strain N199)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: H81703  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: H81703  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-614 <TET>  
A:Cross-references: GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF39278.1; PID:g719  
A:Experimental source: strain N199 (Mopn)  
C:Genetics:  
A:Gene: TC0422

Query Match 85.4%; Score 35; DB 2; Length 614;  
Best Local Similarity 62.5%; Pred. No. 21;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
| | | | | | | |  
Db 453 GRLLVOPG 460

## RESULT 3

T36415

probable iron-siderophore permease transmembrane protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C:Accession: T36415

R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21606

A:Accession: T36415

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-358 &lt;SNU&gt;

A:Cross-references: EMBL:AL109974; PIDN:CAB53327.1; GSPDB:GN00070; SCOEDB:SCF34.16c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCF34.16c

C:Superfamily: ferrichrome ABC transporter

Query Match 82.9%; Score 34; DB 2; Length 358;  
Best Local Similarity 75.0%; Pred. No. 20;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
| | | | | | | |  
Db 323 GRLLVOPG 330

## RESULT 4

DB3347

probable aminotransferase PA2394 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: DB3347

R:Stover, C.K.; Plam, X.Q.; Eryin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337

A:Accession: DB3347

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-427 &lt;STO&gt;

A:Cross-references: GB:AE004665; GB:AE004091; NID:g9948426; PIDN:AG05782.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2394

Query Match 82.9%; Score 34; DB 2; Length 427;  
Best Local Similarity 87.5%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
| | | | | | | |  
Db 131 GRLLVOPG 138

## RESULT 5

T35549

hypothetical protein SC63.05 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C:Accession: T35549

C:Accession: T35549

R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999

A:Reference number: Z21582

A:Accession: T35549

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-588 &lt;SEED&gt;

A:Cross-references: EMBL:AL096825; PIDN:CAB46977.1; GSPDB:GN00070; SCOEDB:SC63.05

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC63.05

Query Match 82.9%; Score 34; DB 2; Length 588;  
Best Local Similarity 75.0%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
| | | | | | | |  
Db 320 GRLLVOPG 327

## RESULT 6

DB1556

conserved hypothetical protein CP0625 [imported] - Chlamydia pneumoniae (strain A

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C:Accession: DB1556

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39

A:Reference number: AB1500; MUID:20150255

A:Accession: DB1556

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-619 &lt;REA&gt;

A:Cross-references: GB:AE002220; GB:AE002161; NID:g7189536; PIDN:AAF38440.1; PID:g718

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0625

Query Match 82.9%; Score 34; DB 2; Length 619;  
Best Local Similarity 62.5%; Pred. No. 34;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOPG 8  
| | | | | | | |  
Db 458 GRLLVOPG 465

## RESULT 7

A72114

s/t protein kinase - Chlamydia pneumoniae (strain CWL029)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C:Accession: A72114

R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: A72114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-619 &lt;ARN&gt;

A:Cross-references: GB:AE001601; GB:AE001363; NID:g4376409; PIDN:AMD18301.1; PID:g437

A:Experimental source: strain CWL029

C:Genetics:

A:Gene: CPN0148

Query Match 82.9%; Score 34; DB 2; Length 619;

Best Local Similarity 62.5%; Pred. No. 34;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GRLVOPG 8  
||||:||||  
Db 458 GRLVIEPG 465

RESULT 8

T29058  
probable transcription regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T29058

R:Reichenbach, M.; Kleser, H.M.; Denapalte, D.; Elchner, A.; Cullum, J.; Kinashl, H.; Hopw

Mol. Microbiol. 21, 77-96, 1996

A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S

A:Reference number: 220556; MUID:97000351

A:Accession: T29058

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-200 <RED>

A:Cross-references: EMBL:AL031124; PIDN:CAA19976.1

C:Genetics:

A:Note: SCIC2.09

Query Match 80.5%; Score 33; DB 2; Length 200;

Best Local Similarity 75.0%; Pred. No. 18;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLVOPG 8  
||||:||||  
Db 74 GRLTRPG 81

RESULT 9

GB3014  
probable nuclease PA5048 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: GB3014

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lm,

; Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337

A:Accession: GB3014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-255 <STO>

A:Cross-references: GB:AE004918; GB:AE004091; NID:g9951336; PIDN:AA08433.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5048

Query Match 80.5%; Score 33; DB 2; Length 255;

Best Local Similarity 75.0%; Pred. No. 23;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLVOPG 8  
||||:||||  
Db 84 GRLALPG 91

RESULT 10

T17319  
hypothetical protein DKFZp434N178.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T17319

R:Ottenwelder, B.; Obermaier, B.; Meves, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18726

A:Accession: T17319

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-268 <OTF>

A:Cross-references: EMBL:AL117589

A:Experimental source: adult testis; clone DKFZp434N178

C:Genetics:

A:Note: DKFZp434N178.1

Query Match 80.5%; Score 33; DB 2; Length 268;  
Best Local Similarity 62.5%; Pred. No. 24;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GRLVOPG 8  
||||:||||  
Db 196 GRLMEPG 203

RESULT 11

F83166  
hypothetical protein PA3825 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: F83166

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

; Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: AB2950; MUID:20437337

A:Accession: F83166

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-526 <STO>

A:Cross-references: GB:AE004800; GB:AE004091; NID:g9949999; PIDN:AA07212.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3825

C:Superfamily: probable membrane protein y1ab

Query Match 80.5%; Score 33; DB 2; Length 526;

Best Local Similarity 75.0%; Pred. No. 47;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLVOPG 8  
||||:||||  
Db 127 GRLMLPG 134

RESULT 12

A43317  
germ cell-less protein gcl - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998

C:Accession: A43317

R:Jongens, T.A.; Hay, B.; Jan, L.Y.; Jan, Y.N.

Cell 70, 569-584, 1992

A:Title: The germ cell-less gene product: a posteriorly localized component necessary

A:Reference number: A43317; MUID:92370679

A:Accession: A43317

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-569 <CON>

A:Cross-references: GB:M97933; NID:g157489; PID:g157490

A:Note: sequence extracted from NCBI backbone (NCBIN:111435, NCBIP:111440)

C:Genetics:

A:Gene: FlyBase:gcl

A:Cross-references: FlyBase:FBgn0005695

Query Match 80.5%; Score 33; DB 2; Length 569;  
Best Local Similarity 62.5%; Pred. No. 51;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVPG 8  
|:|:|:|  
Db 379 GRMLLEPG 386

## RESULT 13

B44212  
structural protein 2 precursor - hepatitis E virus (strain Mexico)  
C/Species: hepatitis E virus  
C/Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999  
C/Accession: B44212  
R/Huang, C.C.; Nguyen, D.; Fernandez, J.; Yun, K.Y.; Fry, K.E.; Bradley, D.W.; Tam, A.W.  
Virology 191, 550-558, 1992  
A/Title: Molecular cloning and sequencing of the Mexico isolate of hepatitis E virus (HEV)  
A/Reference number: A44212; MUID:93079857  
A/Accession: B44212  
A/Molecule type: genomic RNA  
A/Residues: 1-659 <HDA>  
A/Cross-references: GB:M74506; NID:g3330017; PIDN:AAA45732.1; PID:g3330020  
C/Superfamily: hepatitis E virus structural protein 2  
C/Keywords: structural protein  
F:1.22/Domain: signal sequence #status predicted <SIG>  
F:23-659/Product: structural protein 2 #status predicted <SP2>

Query Match 80.5%; Score 33; DB 1; Length 659;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLIVPG 8  
|:|:|:|  
Db 233 RLIVPG 239

## RESULT 14

VHMMH2  
structural protein 2 precursor - hepatitis E virus (strain Burma)  
C/Species: hepatitis E virus  
C/Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1999  
C/Accession: C40778  
R/Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.  
Virology 185, 120-131, 1991  
A/Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome  
A/Reference number: A40778; MUID:92024067  
A/Accession: C40778  
A/Molecule type: genomic RNA  
A/Residues: 1-660 <TRAM>  
A/Cross-references: GB:M73218; NID:g3330023; PIDN:AAA45736.1; PID:g3330026  
A/Note: the authors translated the codon GGC for residue 2 as Ala  
C/Superfamily: hepatitis E virus structural protein 2  
C/Keywords: structural protein  
F:1.22/Domain: signal sequence #status predicted <SIG>  
F:23-660/Product: structural protein 2 #status predicted <SP2>

Query Match 80.5%; Score 33; DB 1; Length 660;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLIVPG 8  
|:|:|:|  
Db 233 RLIVPG 239

RESULT 15  
T30517  
complement C3-Q2 - common carp

C/Species: Cyprinus carpio (common carp)  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C/Accession: T30517  
R/Nakao, M.; Obo, R.; Mutsuro, J.; Fujiki, K.; Nonaka, M.; Yano, T.  
submitted to the EMBL Data Library, July 1998  
A/Description: Sequence diversity of the third complement component (C3) from a bony fish  
A/Reference number: Z20840  
A/Accession: T30517

A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1230 <NAK>  
A/Cross-references: EMBL:AB016215; NID:d1253045; PID:d1037610; PIDN:BAA36623.1

Query Match 80.5%; Score 33; DB 2; Length 1230;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVPG 7  
|:|:|:|  
Db 572 GRLLVPG 578

Search completed: June 13, 2001, 14:10:41  
Job time: 142 sec



**This Page Blank (uspto)**

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:42 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-7

Sequence: 1 GRLVOPG 8

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	461	1 HGD_ARATH	092ra2 arabidopsis
2	35	85.4	1122	1 TERT_MOUSE	070372 mus musculus
3	33	80.5	485	1 VST2_HEYRH	000270 hepatitis e
4	33	80.5	569	1 GCL1_DROME	001820 drosophila
5	33	80.5	659	1 VST2_HEYME	003500 hepatitis e
6	33	80.5	660	1 VST2_HEYBU	P29336 hepatitis e
7	33	80.5	660	1 VST2_HEYMY	004611 hepatitis e
8	33	80.5	660	1 VST2_HEYPA	P33426 hepatitis e
9	33	80.5	713	1 GAC1_HUMAN	075325 homo sapien
10	31	75.6	132	1 Y055_ARCFU	030181 archaeoglob
11	31	75.6	136	1 CAIE_ECOLI	P39206 escherichia
12	31	75.6	286	1 FPG_STRCO	092b26 streptomyce
13	31	75.6	336	1 EFTU_MICLU	P09953 micrococcu
14	31	75.6	403	1 CHEB_TREPA	083639 streptococ
15	31	75.6	448	1 HGD_EMEHI	000657 emericella
16	31	75.6	539	1 OGP_SHEEP	028542 ovis aries
17	31	75.6	836	1 GLK1_MOUSE	060934 mus musculu
18	31	75.6	918	1 GLK1_HUMAN	P39086 homo sapien
19	31	75.6	1875	1 ITB4_HUMAN	P16144 homo sapien
20	30	73.2	85	1 PEM1_ECOLI	P13975 escherichia
21	30	73.2	103	1 ERH_AEDAE	093104 aedes aegypt
22	30	73.2	104	1 ER_DROME	024337 drosophila
23	30	73.2	104	1 ER_DROME	094554 drosophila
24	30	73.2	171	1 YSNB_BACSU	P94559 bacillus su
25	30	73.2	367	1 INX2_DROME	09v427 drosophila
26	30	73.2	394	1 EFTU_UREPA	P50068 ureaplasma
27	30	73.2	405	1 EFTU_AOUAE	056429 aquifex aeo
28	30	73.2	405	1 EFTU_AOUAE	050283 aquifex pyr
29	30	73.2	445	1 HGD_HUMAN	093039 homo sapien
30	30	73.2	460	1 YMT2_CAEEL	P34479 caenorhabdi
31	30	73.2	471	1 IAAG_MAIZE	Q41819 zea mays (m
32	30	73.2	481	1 SYM_MOUSE	P32921 mus musculu
33	30	73.2	520	1 YM23_MYCTU	Q10508 mycobacteri

34	30	73.2	564	1 DPV4_RAT	062951 rattus norv
35	30	73.2	572	1 DPV4_MOUSE	035098 mus musculu
36	30	73.2	652	1 NOS2_PARDE	051705 paracoccus
37	30	73.2	952	1 YH19_RHOCA	002998 rhodobacter
38	30	73.2	996	1 PIP2_YEAST	P52960 saccharomyc
39	30	73.2	3054	1 POLG_TEV	P04517 t genome po
40	29	70.7	69	1 NS2C_HUMAN	Q14951 homo sapien
41	29	70.7	116	1 HV3Q_HUMAN	P01778 homo sapien
42	29	70.7	167	1 APE2_AERPE	073943 aeropyrum p
43	29	70.7	167	1 FIXH_RHIME	P18397 rhizobium m
44	29	70.7	168	1 IA03_WHEAT	P17314 triticum ae
45	29	70.7	171	1 IAAD_HORVU	P11643 hordeum vul

## ALIGNMENTS

```

RESULT 1
HGD_ARATH          STANDARD:      PRT: 461 AA.
ID HGD_ARATH
AC 092RA2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOMOGENTISATE 1,2-DIOXYGENASE (EC 1.13.11.5) (HOMOGENTISICASE)
DE (HOMOGENTISATE OXYGENASE) (HOMOGENTISIC ACID OXIDASE).
GN HGO.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid:3702.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Schmidt S.R., Werner E., Mueller C.R., Kress W.;
RT "Cloning and characterization of the homogentisate 1,2-dioxygenase
RT gene in A. thaliana and C. elegans.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HOMOGENTISATE + O(2) -> 4-MALEYLACETACETATE.
CC -1- COFACTOR: IRON.
CC -1- PATHWAY: CATABOLISM OF TYROSINE; THIRD STEP, CATABOLISM OF
CC PHENYLALANINE; FOURTH STEP.
CC -1- SIMILARITY: BELONGS TO THE HOMOGENTISATE DIOXYGENASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U80668; AAD00360.1; -
DR Oxioreductase; Dioxigenase; Iron; Phenylalanine catabolism;
KW Tyrosine catabolism.
SQ
SEQUENCE 461 AA; 51469 MW; CD91768C88014D31 CRC64;

Query Match          85.4%; Score 35; DB 1; Length 461;
Best Local Similarity 87.5%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GRLVOPG 8
Db 176 GRLVTPG 183

RESULT 2
HGD_ARATH          STANDARD:      PRT: 1122 AA.
ID TERT_MOUSE
AC 070372; 035432;
DT 30-MAY-2000 (Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TELOMERASE REVERSE TRANSCRIPTASE (EC 2.7.7.-) (TELOMERASE CATALYTIC  
 DE SUBUNIT).  
 GN TERT.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98241176; PubMed=9582020;  
 RA Greenberg R.A., Allsopp R.C., Chin L., Morin G.B., Depinho R.A.;  
 RT "Expression of mouse telomerase reverse transcriptase during  
 development, differentiation and proliferation.";  
 RL Oncogene 16:1723-1730(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98393668; PubMed=9724727;  
 RA Martin-Rivera L., Herrera E., Albar J.P., Blasco M.A.;  
 RT "Expression of mouse telomerase catalytic subunit in embryos and  
 adult tissues.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10471-10476(1998).  
 RN [3]  
 RP SEQUENCE OF 550-616 FROM N.A.  
 RA Drissi R., Cleveland J.L.;  
 RT "Partial sequence of Mus musculus telomerase catalytic subunit  
 homolog.";  
 RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR  
 THE REPLICATION OF CHROMOSOME TERRINI IN MOST EUKARYOTES. IT  
 ELONGATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT ADDS  
 SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE  
 SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.  
 CC TELOMERASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF051911; AAC09323.1; -  
 DR EMBL, AF073311; AAC34821.1; -  
 DR EMBL, AF029235; AAB84200.1; -  
 DR MGI:1202709; TERT.  
 KW Transferrase; RNA-directed DNA polymerase; Telomere; Nuclear protein;  
 KW DNA-binding.  
 FT CONFLICT 553 553 I -> V (IN REF. 3).  
 FT SEQUENCE 1122 AA; 127977 MW; F8526905DD558C CRC64;  
 SQ  
 Query Match 85.4%; Score 35; DB 1; Length 1122;  
 Best Local Similarity 87.5%; Pred. NO. 14;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GRLVOPG 8  
 Db 35 GRLVOPG 42.  
 RESULT 3  
 VST2\_HEVRH STANDARD; PRT; 485 AA.  
 ID VST2\_HEVRH  
 AC 000270;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE STRUCTURAL PROTEIN 2 (FRAGMENT).

OS Hepatitis E virus (Isolate Rhesus) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.  
 OX NCBI\_TaxID=31766;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92261377; PubMed=1584074;  
 RA Uchida T., Suzuki K., Hayashi N., Iida F., Hara T., Oo S.S.,  
 RA Wang C.-K., Shikata T., Ichikawa M., Rishisa T., Mizuno K.,  
 RA Min K.M.;  
 RT "Hepatitis E virus: cDNA cloning and expression.";  
 RL Microbiol. Immunol. 36:67-79(1992).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D90274; BAA20910.1; -  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT SEQUENCE 485 AA; 52317 MW; 5A4FD9273AC74F94 CRC64;  
 SQ  
 Query Match 80.5%; Score 33; DB 1; Length 485;  
 Best Local Similarity 85.7%; Pred. NO. 16;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2 RLTVOPG 8  
 Db 101 RLTVOPG 107

RESULT 4  
 GCL1\_DROME STANDARD; PRT; 569 AA.  
 ID GCL1\_DROME  
 AC 001820; G9V4X5;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GERM CELL-LESS PROTEIN.  
 GN GCL OR CG8411.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 DE Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92370679; PubMed=1380406;  
 RA Jorgens T.A., Hay B., Jan L.Y., Jan Y.N.;  
 RT "The germ cell-less gene product: a posteriorly localized component  
 necessary for germ cell development in Drosophila.";  
 RL Cell 70:569-584(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RC STRAIN-BERKELEY;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gccayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Chame M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Adayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gebhart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasek P., Lei Y., Levitsky A.A., Li Z., Li Z., Liang Y., Lin X.,  
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Modyarty C., Morris J., Mosher A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reibert C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spleit E., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: REQUIRED FOR THE SPECIFICATION OF POLE CELLS AND GERM  
CC CELL FORMATION. MOTHERS WITH REDUCED GLC FUNCTION GIVE RISE TO  
CC STERILE ADULT PROGENY THAT LACK GERM CELLS.  
CC -1- SUBCELLULAR LOCATION: POLE PLASM.  
CC -1- DEVELOPMENTAL STAGE: OOCYTESIS AND EMBRYOGENESIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M97933; AAA28566.1; -  
CC EMBL: AE003836; AAF59048.1; -  
CC PIR: A43317; A43317.  
CC DR Flybase: FBgn0005695; gcl.  
CC DR InterPro: IPR000210; -  
CC DR Pfam: PF00651; BTB.1.  
CC DR PROSITE: PS50097; BTB.1.  
CC KW Developmental protein.  
CC SQ SEQUENCE 569 AA; 65152 MW; A91A670C4B5F83A4 CRC64;  
  
Query Match 80.5%; Score 33; DB 1; Length 569;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRLLVPG 8  
Db 379 GRMLLEPG 386  
  
RESULT 5  
VST2\_HEVME STANDARD; PRT; 659 AA.  
AC 003500;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE STRUCTURAL PROTEIN 2 PRECURSOR.  
OS Hepatitis E virus (strain Mexico) (HEV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage.  
OX NCBI\_TaxID=31768;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93079857; PubMed=1448913;

RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,  
RA Bradley D.W., Tam A.W., Reyes G.R.;  
RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis  
RT E virus (HEV)."  
RL Virology 191:550-558(1992).  
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M73218; AAA45736.1; -  
CC PIR: B44212; B44212.  
CC DR Signal.  
CC DR PIR: B44212; B44212.  
CC KW Signal.  
CC FT SIGNAL 1 22 BY SIMILARITY.  
CC FT CHAIN 23 659 STRUCTURAL PROTEIN 2.  
CC SQ SEQUENCE 659 AA; 70640 MW; CF75E75EFD8FBE2C CRC64;  
  
Query Match 80.5%; Score 33; DB 1; Length 659;  
Best Local Similarity 85.7%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 RLTVQPG 8  
Db 233 RLTVQPG 239  
  
RESULT 6  
VST2\_HEVBU STANDARD; PRT; 660 AA.  
AC P29326;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE STRUCTURAL PROTEIN 2 PRECURSOR (ORF2).  
OS Hepatitis E virus (strain Burma) (HEV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage.  
OX NCBI\_TaxID=31767;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92024067; PubMed=1926770;  
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,  
RA Fry K.E., Reyes G.R.;  
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the  
RT full-length viral genome."  
RL Virology 185:120-131(1991).  
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M73218; AAA45736.1; -  
CC PIR: C40778; VHWK2.  
CC DR Signal.  
CC DR SIGNAL 1 19 POTENTIAL.  
CC FT CHAIN 20 660 STRUCTURAL PROTEIN 2.  
CC SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCCA461C CRC64;

Query Match 80.5%; Score 33; DB 1; Length 660;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
 1:|||||  
 Db 233 RLVOPG 239

RESULT 7  
 VST2.HEVPA STANDARD; PRT; 660 AA.  
 AC 004611;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE STRUCTURAL PROTEIN 2 PRECURSOR (ORF2).  
 OS Hepatitis E virus (strain Myanmar) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.  
 OX NCBI\_TaxID=31769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93227573; PubMed=8470371;  
 RA Ave T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,  
 RA Rikhisia T., Winn K.;  
 RT "Sequence and gene structure of the hepatitis E virus isolated from Myanmar."  
 RL Virus Genes 7:95-109(1993).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10330; BAA01174.1; -  
 DR Signal.  
 KW SIGNAL.  
 FT CHAIN  
 FT SIGNAL  
 SQ SEQUENCE 660 AA; 70998 MW; 3A82A4EA25C6253 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 660;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
 1:|||||  
 Db 233 RLVOPG 239

RESULT 8  
 VST2.HEVPA STANDARD; PRT; 660 AA.  
 AC P33426;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE STRUCTURAL PROTEIN 2 PRECURSOR (ORF2).  
 OS Hepatitis E virus (strain Pakistan) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.  
 OX NCBI\_TaxID=33774;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92115700; PubMed=1733327;  
 RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,  
 RA Malik I.A., Iqbal M., Purcell R.H.;  
 RT "Characterization of a prototype strain of hepatitis E virus,"

RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M80581; AAA45727.1; -  
 DR Signal.  
 KW SIGNAL.  
 FT CHAIN  
 FT SIGNAL  
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 660;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
 1:|||||  
 Db 233 RLVOPG 239

RESULT 9  
 GAC1.HUMAN STANDARD; PRT; 713 AA.  
 AC 075325;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GLIOMA AMPLIFIED ON CHROMOSOME 1 PROTEIN PRECURSOR.  
 GN GAC1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Glial tumor;  
 RX MEDLINE=98324709; PubMed=9662332;  
 RA Malfroy B., Almeida A., Zhu X.X., Vogt N., Tyagi R., Mulleris M.,  
 RA Dutrillaux A.-M., Dutrillaux B., Ross D., Hanash S.;  
 RT "GAC1, a new member of the leucine-rich repeat superfamily on  
 chromosome band 1q32.1, is amplified and overexpressed in malignant  
 gliomas."  
 RL Oncogene 16:2997-3002(1998).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: OVERAMPLIFIED IN MALIGNANT GLIOMAS.  
 CC -1- SIMILARITY: CONTAINS 11 LEUCINE-RICH REPEATS (LRR).  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF030435; AAC39792.1; -  
 DR MIM: 605492;  
 DR InterPro: IPR000372; -  
 DR InterPro: IPR000483; -  
 DR InterPro: IPR001611; -  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; Ig; 1.  
 DR Pfam: PF00560; LRR; 9.

DR Pfam: PF01462; LRRT: 1.  
DR Pfam: PF01463; LRRT: 1.  
DR PRINTS: PR00019; LEURICHP.  
KW Immunoglobulin domain; Transmembrane; Glycoprotein; Repeat;  
KW Leucine-rich repeat; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 73  
FT DOMAIN 19 630  
FT TRANSMEM 631 651  
FT DOMAIN 652 713  
FT REPEAT 92 115  
FT REPEAT 116 139  
FT REPEAT 140 163  
FT REPEAT 165 187  
FT REPEAT 188 211  
FT REPEAT 213 235  
FT REPEAT 236 259  
FT REPEAT 261 283  
FT REPEAT 309 333  
FT REPEAT 334 357  
FT REPEAT 359 385  
FT DOMAIN 438 504  
FT DISULFID 445 497  
FT CARBOHYD 94 94  
FT CARBOHYD 381 381  
FT CARBOHYD 555 555  
FT CARBOHYD 583 583  
SQ SEQUENCE 713 AA; 78798 MW; EC8BC0FD240C9396 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 713;  
Best Local Similarity 75.0%; Pred. No. 23;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLLVOPG 8  
Db 515 GRLLVOPG 522

RESULT 10  
ID Y055\_ARCFU STANDARD; PRT; 132 AA.  
AC 030181;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL PROTEIN AF0055.  
GN AF0055.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RC MEDLINE=98049343; PubMed=9389475;  
RA Kient H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Krelavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Klrness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Arlisch P., Kaine B.P., Sykes S.M.,  
RA Mason P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -1- SIMILARITY: BELONGS TO THE UPF0107 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE001102; MB91166.1; -.  
DR TIGR: AF0055; -.  
DR InterPro: IPR002840; -.  
DR Pfam: PF01989; DUF126; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 132 AA; 14189 MW; 9C97ABCF9512F20D CRC64;

Query Match 75.6%; Score 31; DB 1; Length 132;  
Best Local Similarity 75.0%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLLVOPG 8  
Db 50 GRLLVOPG 57

RESULT 11  
ID CAIE\_ECOLI STANDARD; PRT; 196 AA.  
AC P39206;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CARNITINE OPERON PROTEIN CAIE.  
GN CAIE.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O44 K74;  
RC MEDLINE=95115548; PubMed=7815937;  
RA Eichler K., Bourgeois F., Buchet A., Kleber H.-P.,  
RA Mandrand-Berthelot M.-A.;  
RT "Molecular characterization of the cai operon necessary for carnitine  
RT metabolism in Escherichia coli.";  
RL Mol. Microbiol. 13:775-786(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RC MEDLINE=92334977; PubMed=1630901;  
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
RA Isono K., Mizobuchi K., Nakata A.;  
RT "Systematic sequencing of the Escherichia coli genome: analysis of  
RT the 0-2.4 min region.";  
RL Nucleic Acids Res. 20:3305-3308(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RC MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -1- PATHWAY: CARNITINE METABOLISM.  
CC -1- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NDL FAMILY OF  
CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC



ID CHEB TREPA STANDARD: PRT: 403 AA.  
 AC 083639;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE PROTEIN-GLUTAMATE METHYLESTERASE (EC 3.1.1.61).  
 GN CHEB OR TP0631.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RX NCBI\_Taxid=160;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLS.  
 RA MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.R., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete."  
 RL Science 281:375-388(1998).  
 CC -1- FUNCTION: INVOLVED IN THE MODULATION OF THE CHEMOTAXIS SYSTEM;  
 CATALYZES THE DEMETHYLATION OF SPECIFIC METHYLGLUTAMATE RESIDUES  
 INTRODUCED INTO THE CHEMORECEPTORS (METHYL-ACCEPTING CHEMOTAXIS  
 PROTEINS) BY CHER (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: PROTEIN L-GLUTAMATE O-METHYL ESTER + H(2)O =  
 CC PROTEIN L-GLUTAMATE + METHANOL.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- DOMAIN: THE N-TERMINAL REGULATORY DOMAIN INHIBITS THE ACTIVITY OF  
 CC THE C-TERMINAL EFFECTOR DOMAIN.  
 CC -1- PTM: PHOSPHORYLATED BY CHEA. PHOSPHORYLATION SUPPRESSES THE  
 CC INHIBITORY ACTIVITY OF THE N-TERMINAL DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE RESPONSE  
 CC REGULATORY FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE CHEB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE001238; AAC65606.1; -  
 DR HSSP: P04042; 1CHD.  
 DR TIGR: TP0631; -  
 DR InterPro: IPR000673; -  
 DR InterPro: IPR001789; -  
 DR Pfam: PF01339; Cheb\_methylase; 1.  
 DR Pfam: PF00072; response\_reg; 1.  
 KW Hydrolyase; Chemotaxis; Sensory transduction; Phosphorylation.  
 FT DOMAIN 1 138 RESPONSE REGULATORY DOMAIN.  
 FT DOMAIN 139 202 LINKER.  
 FT DOMAIN 203 403 PROTEIN-GLUTAMATE METHYLESTERASE.  
 FT MOD\_RES 59 59 PHOSPHORYLATION (BY SIMILARITY).  
 FT ACT\_SITE 219 219 BY SIMILARITY.  
 FT ACT\_SITE 246 246 BY SIMILARITY.  
 FT ACT\_SITE 342 342 BY SIMILARITY.  
 SQ SEQUENCE 403 AA; 44098 MW; E629B147BF3AD03A CRC64;

Query Match 75.6%; Score 31; DB 1; Length 403;  
 Best local Similarity 62.5%; Pred. No. 35;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVOPG 8  
 DB 279 GRVLIAPG 286

RESULT 15  
 ID HGD\_EMENT STANDARD: PRT: 448 AA.  
 AC 000667;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE HOMOGENTISATE 1,2-DIOXYGENASE (EC 1.13.11.5) (HOMOGENTISICASE)  
 DE (HOMOGENTISATE OXYGENASE) (HOMOGENTISIC ACID OXIDASE).  
 GN HMG.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiaceae; Trichocomaceae; Emericella.  
 RX NCBI\_Taxid=5072;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BI-1.  
 RA MEDLINE=95403355; PubMed=7673153;  
 RA Fernandez-Canon J.M., Penalba M.A.;  
 RT "Molecular characterization of a gene encoding a homogenisate  
 RT dioxygenase from Aspergillus nidulans and identification of its human  
 RT and plant homologues."  
 RL J. Biol. Chem. 270:21199-21205(1995).  
 CC -1- CATALYTIC ACTIVITY: HOMOGENTISATE + O(2) = 4-MALEYLACETATE.  
 CC -1- COFACTOR: IRON.  
 CC -1- PATHWAY: CATABOLISM OF TYROSINE; THIRD STEP, CATABOLISM OF  
 CC PHENYLALANINE; FOURTH STEP. ALSO INVOLVED IN PHENYLACETATE  
 CC CATABOLISM.  
 CC -1- INDUCTION: DURING GROWTH WITH PHENYLACETATE AND PHENYLALANINE.  
 CC -1- SIMILARITY: BELONGS TO THE HOMOGENTISATE DIOXYGENASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U30797; AAC49071.1; -  
 DR EMBL: AJ001836; CA05042.1; -  
 KW Oxidoreductase; Dioxygenase; Iron; Phenylalanine catabolism;  
 KW Tyrosine catabolism.  
 SQ SEQUENCE 448 AA; 50168 MW; 4AB6414823A04COD CRC64;

Query Match 75.6%; Score 31; DB 1; Length 448;  
 Best local Similarity 85.7%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVOP 7  
 DB 174 GRLLVOP 180

Search completed: June 13, 2001, 14:21:42  
 Job time: 802 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:26 ; Search time 225.85 Seconds

(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-7

Perfect score: 41

Sequence: 1 GRLLVPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.15:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	97.6	547	4	095072
2	40	97.6	547	4	095072
3	35	85.4	104	11	09JLMI
4	35	85.4	461	10	09LDB8
5	35	85.4	514	11	09R266
6	35	85.4	614	2	084147
7	35	85.4	614	2	09PKP3
8	35	85.4	1128	11	09QXK4
9	34	82.9	358	2	09RK09
10	34	82.9	399	5	09NGF2
11	34	82.9	399	5	09N6A6
12	34	82.9	588	2	09XAH5
13	34	82.9	619	2	09Z935
14	34	82.9	619	2	09K228
15	34	82.9	2189	5	09VXD5
16	33	80.5	200	2	086510
17	33	80.5	227	14	056048
18	33	80.5	227	14	056049
19	33	80.5	227	14	09W8Z7

20	33	80.5	268	4	09UEF3	09ufl3 homo sapien
21	33	80.5	283	14	09YBP1	09ypb1 hepatitis e
22	33	80.5	436	14	09W148	09w148 hepatitis e
23	33	80.5	466	14	09WNN1	09wnn1 hepatitis e
24	33	80.5	486	14	09WNN2	09wnn2 hepatitis e
25	33	80.5	525	14	039947	039947 hepatitis e
26	33	80.5	603	13	09PT11	09pt11 xenopus lae
27	33	80.5	605	14	09YWL0	09ywl0 hepatitis e
28	33	80.5	660	14	069419	069419 hepatitis e
29	33	80.5	660	14	069411	069411 hepatitis e
30	33	80.5	660	14	081871	081871 hepatitis e
31	33	80.5	660	14	089468	089468 hepatitis e
32	33	80.5	660	14	071147	071147 hepatitis e
33	33	80.5	660	14	068985	068985 hepatitis e
34	33	80.5	660	14	091855	091855 hepatitis e
35	33	80.5	660	14	091856	091856 hepatitis e
36	33	80.5	660	14	081878	081878 hepatitis e
37	33	80.5	660	14	036613	036613 swine hepat
38	33	80.5	660	14	09YLR2	09ylr2 hepatitis e
39	33	80.5	660	14	09YLO9	09ylo9 hepatitis e
40	33	80.5	660	14	09WQAO	09wqao hepatitis e
41	33	80.5	660	14	09WLL4	09wll4 hepatitis e
42	33	80.5	660	14	09W7W9	09w7w9 hepatitis e
43	33	80.5	672	14	091VZ8	091vz8 hepatitis e
44	33	80.5	713	4	075325	075325 homo sapien
45	33	80.5	1230	13	09YIAS	09yias cyprinus ca

## ALIGNMENTS

RESULT 1	095072	PRELIMINARY:	PRT;	547 AA.
AC	095072:			
ID	095072:			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)			
DE	HHR21SPB (FRAGMENT).			
GN	HHR21SPB.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RA	McKay M.J., van der Spek P., Kanaar R., Bootsma D., Hoelmakers J.H.;			
RT	"HHR21SPB, a second human gene homologous to the rrd21			
RT	Schizosaccharomyces pombe DNA double strand break repair gene."			
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF006264; AAD01193.1;			
FT	NON_TER			
FT	547			
SO	SEQUENCE	547 AA;	62613 MW;	544AA31F2D038D7B CRC64;
Query Match	97.6%;	Score 40;	DB 4;	Length 547;
Best Local Similarity	87.5%;	Pred. No. 4.5;		
Matches	7;	Conservative	1;	Mismatches 0;
			Indels	0;
			Gaps	0;
QY	1 GRLLVPG 8			
DB	536 GRLLVPG 543			
RESULT 2	09NVO9	PRELIMINARY:	PRT;	547 AA.
ID	09NVO9:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	CDNA FLJ10573 FIS. CLONE NT2RP2003177.			

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NEO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF001435; BAA91690.1; -
SO SEQUENCE 547 AA; 62585 MW; 05A1CFC874F6C3BB CRC64;

Query Match
Best Local Similarity 97.6%; Score 40; DB 4; Length 547;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVOPG 8
Db 536 GRLLVOPG 543

RESULT 3
Q9JLM1 PRELIMINARY; PRT; 104 AA.
AC Q9JLM1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE TLOMERASE REVERSE TRANSCRIPTASE (FRAGMENT).
GN TERT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hatakeyama S., Ishikawa F.;
RT "Identification of the mouse telomerase reverse transcriptase (mTERT)
RT promoter."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF157502; AAF42984.1; -
KM RNA-directed DNA polymerase.
FT NON_TER 104
SQ SEQUENCE 104 AA; 11998 MW; 4B649E63476D3D44 CRC64;

Query Match
Best Local Similarity 85.4%; Score 35; DB 11; Length 104;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVOPG 8
Db 35 GRLLVOPG 42

RESULT 4
Q9LDB8 PRELIMINARY; PRT; 461 AA.
AC Q9LDB8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HOMOGENTISATE 1,2-DIOXYGENASE (EC 1.13.11.5).
GN HGO.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.

```

```

OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Schmidt S.R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB018115; BAA97130.1; -
DR EMBL: AF130845; AAF36499.1; -
KW Dioxynase; Oxidoreductase.
SQ SEQUENCE 461 AA; 51455 MW; CD90DD9638014D31 CRC64;

Query Match
Best Local Similarity 85.4%; Score 35; DB 10; Length 461;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVOPG 8
Db 176 GRLLVOPG 183

RESULT 5
Q9R266 PRELIMINARY; PRT; 514 AA.
AC Q9R266;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE TLOMERASE REVERSE TRANSCRIPTASE (FRAGMENT).
GN TERT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99144726; PubMed=10022128;
RA Greenberg R.A., O'Hagan R.C., Deng H., Xiao Q., Hann S.R., Adams R.R.,
RA Lichtsteiner S., Chin L., Morin G.B., DePinho R.A.;
RT "Telomerase reverse transcriptase gene is a direct target of c-Myc but
RT is not functionally equivalent in cellular transformation."
RL Oncogene 18:1219-1226(1999).
DR EMBL: AF121949; AAD24465.1; -
KW RNA-directed DNA polymerase.
FT NON_TER 514
SQ SEQUENCE 514 AA; 58382 MW; 3096599776D9B8FD CRC64;

Query Match
Best Local Similarity 85.4%; Score 35; DB 11; Length 514;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVOPG 8
Db 35 GRLLVOPG 42

RESULT 6
O84147 PRELIMINARY; PRT; 614 AA.
AC O84147;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE SERINE/THREONINE PROTEIN KINASE.

```

GN CT145.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/UV-3/CX;  
 RA MEDLINE=99000809; PubMed=9784136;  
 RX Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis."  
 RL Science 282:754-759(1998).  
 DR EMBL: AE001288; AAC67736.1; -;  
 DR INTERPRO: IPR000719; -;  
 DR PFAM: PF00069; pkinase; 2.  
 DR PROSITE: PSS0011; PROTEIN\_KINASE\_DOM; 1.  
 KW Serine/threonine-protein kinase.  
 SO SEQUENCE 614 AA; 69693 MW; 28BEC559CD694001 CRC64;

Query Match 85.4%; Score 35; DB 2; Length 614;  
 Best Local Similarity 62.5%; Pred. No. 54;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRLVQPG 8  
 11111111  
 Db 453 GRLIEPG 460

RESULT 7  
 O9PKP3 PRELIMINARY; PRT; 614 AA.  
 AC O9PKP3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN TC0422.  
 GN TC0422.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MOPN / NIG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
 pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
 RA Hickey E.K., Peterson J., Uitterback T., Berry K.,  
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 RA Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
 RA Salzberg S.L., Eisen J., Fraser C.M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE002309; AAF39278.1; -;  
 DR TIGR: TC0422; -;  
 DR INTERPRO: IPR000719; -;  
 DR PROSITE: PSS0011; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein.  
 SO SEQUENCE 614 AA; 69781 MW; A24D88E70258B75 CRC64;

Query Match 85.4%; Score 35; DB 2; Length 614;  
 Best Local Similarity 62.5%; Pred. No. 54;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRLVQPG 8  
 11111111  
 Db 453 GRLIEPG 460

RESULT 8  
 O9QXZ4 PRELIMINARY; PRT; 1128 AA.  
 AC O9QXZ4;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE TELOMERASE CATALYTIC SUBUNIT.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Guo W., Okamoto M., Baluda M.A., Park N.-H.;  
 RT "Cloning of the cDNA and promoter of hamster telomerase catalytic  
 subunit (hamTERT)."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF149012; AAF1734.1; -;  
 SO SEQUENCE 1128 AA; 128393 MW; 1D4F81249012174E CRC64;

Query Match 85.4%; Score 35; DB 11; Length 1128;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GRLVQPG 8  
 11111111  
 Db 35 GRLVQPG 42

RESULT 9  
 O9RK09 PRELIMINARY; PRT; 358 AA.  
 AC O9RK09;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE PUTATIVE IRON-SIDOPHORE PERMEASE TRANSMEMBRANE PROTEIN.  
 GN SCE34.16c.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1902;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Saunders D.C., Harris D.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL MOL. Microbiol. 21:77-96(1996).  
 DR EMBL: AL109974; CAB53327.1; -;  
 DR INTERPRO: IPR000522; -;  
 DR PFAM: PF01032; RecCD\_family; 1.

KW Transmembrane.  
SQ SEQUENCE 358 AA; 36378 MW; FDD21CD350F11A36 CRC64;

Query Match  
Best Local Similarity 82.9%; Score 34; DB 2; Length 358;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRVLVOPG 8  
|||:||||  
DB 323 GRVLVAPG 330

## RESULT 10

O9NGF2 PRELIMINARY; PRT; 399 AA.

ID O9NGF2  
AC O9NGF2  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE RUDIMENTARY (FRAGMENT).  
GN R.  
OS Drosophila simulans (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7240;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SIM3;  
RA Begun D.J., Whitely P.;  
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).  
DR EMBL; AF252765; AAF68548.1; -.  
FT NON\_TER 1  
FT SEQUENCE 399 AA; 44812 MW; 21ECDA04D1A7CF0 CRC64;

Query Match  
Best Local Similarity 82.9%; Score 34; DB 5; Length 399;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRVLVOPG 8  
|||:||||  
DB 256 GOVLVOPG 263

## RESULT 11

O9N6A6 PRELIMINARY; PRT; 399 AA.

ID O9N6A6  
AC O9N6A6  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE RUDIMENTARY (FRAGMENT).  
GN R.  
OS Drosophila simulans (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7240;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SIM8, SIM4, SIM5, SIM6, AND SIM7;  
RA Begun D.J., Whitely P.;  
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).  
DR EMBL; AF252770; AAF68553.1; -.  
DR EMBL; AF252766; AAF68549.1; -.  
DR EMBL; AF252767; AAF68550.1; -.  
DR EMBL; AF252768; AAF68551.1; -.  
DR EMBL; AF252769; AAF68552.1; -.

FT NON\_TER 1  
FT NON\_TER 399  
SQ SEQUENCE 399 AA; 44759 MW; AAF623F57B1ECCE CRC64;

Query Match  
Best Local Similarity 82.9%; Score 34; DB 5; Length 399;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRVLVOPG 8  
|||:||||  
DB 256 GOVLVOPG 263

## RESULT 12

O9XAH5 PRELIMINARY; PRT; 588 AA.

ID O9XAH5  
AC O9XAH5  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE HYPOTHETICAL 60.3 KDA PROTEIN.  
GN SC63.05.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=9700351; Pubmed=8843436;  
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Mcl. Microbiol. 21:77-96(1996).  
DR EMBL; AL096825; CAB46977.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 588 AA; 60264 MW; A68E50793AA17504 CRC64;

Query Match  
Best Local Similarity 82.9%; Score 34; DB 2; Length 588;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRVLVOPG 8  
|||:||||  
DB 320 GRVLAPG 327

## RESULT 13

O9Z935 PRELIMINARY; PRT; 619 AA.

ID O9Z935  
AC O9Z935  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE S/T PROTEIN KINASE.  
GN CPN0148 OR CPJ0148.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-J138;  
RX MEDLINE=20150255; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RT Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
DR EMBL: AE001601; MAF18301.1; -;  
DR EMBL: AP002545; BAA98358.1; -;  
DR INTERPRO: IPR000719; -;  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Kinase.  
SQ SEQUENCE 619 AA; 70354 MW; CA312E28304D7D80 CRC64;

Query Match 82.9%; Score 34; DB 2; Length 619;  
Best Local Similarity 62.5%; Pred. No. 87;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVPG 8  
DB 458 GRLLVPG 465

RESULT 14  
O9K228  
ID O9K228 PRELIMINARY; PRT; 619 AA.  
AC O9K228;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE HYPOTHETICAL PROTEIN CP0625.  
GN CP0625.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AK39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,  
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,  
RA McCleary G., Salzberg S.L., Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL: AE002220; AAF38440.1; -;  
DR TIGR: CP0625; -;  
SQ SEQUENCE 619 AA; 70326 MW; 2134A8BB204563A2 CRC64;

Query Match 82.9%; Score 34; DB 2; Length 619;  
Best Local Similarity 62.5%; Pred. No. 87;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVPG 8  
DB 458 GRLLVPG 465

RESULT 15  
O9VXD5  
ID O9VXD5 PRELIMINARY; PRT; 2189 AA.  
AC O9VXD5;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE R. GENE PRODUCT.  
GN R.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaniatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
RA Burdick J.M., Busam D.A., Butler H., Caden E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Giodice A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003503; AAF48639.1; -;  
DR HSSP: P00479; JCSU.  
DR FLYBASE: FBgn0003189; r.  
DR INTERPRO: IPR000901; -;  
DR INTERPRO: IPR000901; -;  
DR INTERPRO: IPR001317; -;  
DR INTERPRO: IPR002029; -;  
DR INTERPRO: IPR002082; -;  
DR INTERPRO: IPR002195; -;  
DR INTERPRO: IPR002385; -;  
DR INTERPRO: IPR002474; -;  
DR PFAM: PF00117; GATase; 1.  
DR PFAM: PF00185; OTCase; 1.  
DR PFAM: PF00289; CPSase; 1.  
DR PFAM: PF00744; Dihydroorotase; 1.  
DR PFAM: PF00988; CPSase-sm\_chain; 1.  
DR PRINTS: PR00096; GATase.

DR PRINTS; PR00097; ANTSNTHASEIT.  
DR PRINTS; PR00098; CPSASE.  
DR PRINTS; PR00099; CPSGATASE.  
DR PRINTS; PR00100; AOTCASE.  
DR PRINTS; PR00101; ATCASE.  
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; UNKNOWN\_1.  
DR PROSITE; PS00442; GATASE\_TYPE\_1; 1.  
DR PROSITE; PS00482; DIHYDROCHROTASE\_1; 1.  
DR PROSITE; PS00483; DIHYDROCHROTASE\_2; 1.  
DR PROSITE; PS00866; CPSASE\_1; 1.  
DR PROSITE; PS00867; CPSASE\_2; 2.  
SQ SEQUENCE 2189 AA; 242739 MW; 42FDC09A6136A94F CRC64;

Query Match 82.9%; Score 34; DB 5; Length 2189;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLVOPG 8  
1:|||||  
DB 1777 GQVLVOPG 1784

Search completed: June 13, 2001, 14:20:27  
Job time: 727 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 13, 2001, 14:16:33 : Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-7

Perfect score: 41

Sequence: 1 GRLLVPG 8

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents, AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	87.8	118	2	US-08-652-816A-12
2	33	80.5	246	1	US-07-843-125-11
3	33	80.5	435	1	US-08-259-148A-18
4	33	80.5	435	1	US-08-484-054-18
5	33	80.5	435	2	US-07-876-941A-18
6	33	80.5	436	1	US-08-259-148A-17
7	33	80.5	436	1	US-08-484-054-17
8	33	80.5	436	2	US-07-876-941A-17
9	33	80.5	525	5	PCT-US95-13703-27
10	33	80.5	525	5	PCT-US95-13703-28
11	33	80.5	540	5	PCT-US95-13703-25
12	33	80.5	540	5	PCT-US95-13703-26
13	33	80.5	549	5	PCT-US95-13703-15
14	33	80.5	549	5	PCT-US95-13703-16
15	33	80.5	659	1	US-08-240-049B-16
16	33	80.5	659	1	US-08-259-148A-20
17	33	80.5	659	1	US-08-484-054-20
18	33	80.5	659	2	US-07-876-941A-20
19	33	80.5	660	1	US-08-240-049B-15
20	33	80.5	660	1	US-08-259-148A-19
21	33	80.5	660	2	US-08-484-054-19
22	33	80.5	660	2	US-07-876-941A-19
23	33	80.5	660	3	US-08-840-316-2
24	33	80.5	660	4	US-08-478-507-8
25	33	80.5	660	4	US-08-809-523-2
26	33	80.5	660	5	PCT-US93-08849A-2
27	33	80.5	660	5	PCT-US93-08849-2

## ALIGNMENTS

28	33	80.5	660	5	PCT-US95-13703-13	Sequence 13, Appl
29	33	80.5	660	5	PCT-US95-13703-14	Sequence 14, Appl
30	31	75.6	20	2	US-08-859-931A-4	Sequence 4, Appl
31	31	75.6	44	1	US-08-405-392-11	Sequence 11, Appl
32	31	75.6	44	3	US-08-487-691-11	Sequence 11, Appl
33	31	75.6	100	1	US-08-405-392-12	Sequence 12, Appl
34	31	75.6	100	3	US-08-487-691-12	Sequence 12, Appl
35	31	75.6	554	1	US-08-445-586-7	Sequence 7, Appl
36	31	75.6	556	1	US-08-445-586-2	Sequence 2, Appl
37	31	75.6	849	1	US-08-405-392-18	Sequence 18, Appl
38	31	75.6	849	3	US-08-487-691-18	Sequence 18, Appl
39	31	75.6	866	1	US-08-405-392-17	Sequence 17, Appl
40	31	75.6	866	3	US-08-487-691-17	Sequence 17, Appl
41	31	75.6	905	1	US-08-405-392-2	Sequence 2, Appl
42	31	75.6	905	3	US-08-487-691-2	Sequence 2, Appl
43	31	75.6	905	4	US-08-666-221B-4	Sequence 4, Appl
44	31	75.6	905	4	US-08-666-221B-10	Sequence 10, Appl
45	30	73.2	115	1	US-08-468-661-1	Sequence 1, Appl

RESULT 1  
US-08-652-816A-12  
Sequence 12, Application US/08652816A  
Patent No. 5872215  
GENERAL INFORMATION:  
APPLICANT: Osbourn, JK  
APPLICANT: Allen, DJ  
TITLE OF INVENTION: Specific binding members, materials and  
TITLE OF INVENTION: methods.  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,816A  
FILING DATE: 23-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9125579.4  
FILING DATE: 02-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9125579.8  
FILING DATE: 02-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206372.6  
FILING DATE: 23-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9525004.9  
FILING DATE: 07-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9610824.6  
FILING DATE: 23-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/02240  
FILING DATE: 02-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/244,597  
FILING DATE: 01-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: David W. Clough  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/33308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-652-816A-12

Query Match 87.8%; Score 36; DB 2; Length 118;  
Best Local Similarity 87.5%; Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVPG 8  
| | | | |  
Db 8 GRLLVPG 15

RESULT 2  
US-07-843-125-11  
Sequence 11, Application US/07843125  
Patent No. 5395750  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J  
TITLE OF INVENTION: Methods for producing proteins which  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110-1199  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/843,125  
FILING DATE: 19920228  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Catherine R  
REGISTRATION NUMBER: 34240  
REFERENCE/DOCKET NUMBER: 8594  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-6208  
TELEFAX: (201) 235-3500  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-843-125-11

Query Match 80.5%; Score 33; DB 1; Length 246;  
Best Local Similarity 87.5%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVPG 8  
| | | | |  
Db 8 GRLLVPG 15

RESULT 3  
US-08-259-148A-18  
Sequence 18, Application US/08259148A  
Patent No. 5741490  
GENERAL INFORMATION:  
APPLICANT: Reyes, Gregory R.  
APPLICANT: Bradley, Daniel W.  
APPLICANT: Twu, Jr-Shin  
APPLICANT: Purdy, Michael A.  
APPLICANT: Tam, Albert W.  
APPLICANT: Krawczynski, Krzysztof Z.  
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/259,148A  
FILING DATE: 13-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 822,335  
FILING DATE: 17-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 505,888  
FILING DATE: 05-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 420,921  
FILING DATE: 13-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 367,486  
FILING DATE: 16-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 336,672  
FILING DATE: 11-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 208,997  
FILING DATE: 17-JUN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: C2, MEXICO, FIGURE 9  
US-08-259-148A-18

Query Match 80.5%; Score 33; DB 1; Length 435;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLTVPG 8  
| | | | |



DB 9 RLVOPG 15

RESULT 4  
US-08-484-054-18  
Sequence 18, Application US/08484054  
Patent No. 5770689

## GENERAL INFORMATION:

APPLICANT: Reyes, Gregory R.  
APPLICANT: Bradley, Daniel W.  
APPLICANT: Twu, Jr-Shin  
APPLICANT: Purdy, Michael A.  
APPLICANT: Tam, Albert W.  
APPLICANT: Krawczynski, Krzysztof Z.  
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,054  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 870,985  
FILING DATE: 20-APRIL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 822,335  
FILING DATE: 17-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 505,888  
FILING DATE: 05-APRIL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 420,921  
FILING DATE: 13-OCTOBER-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 367,486  
FILING DATE: 16-JUNE-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 336,672  
FILING DATE: 11-APRIL-1989

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 208,997  
FILING DATE: 17-JUNE-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.38  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: C2, MEXICO, FIGURE 9  
US-08-484-054-18

Query Match

80.5%; Score 33; DB 1; Length 435;

Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;QY 2 RLVOPG 8  
DB 9 RLVOPG 15RESULT 5  
US-07-876-941A-18  
Sequence 18, Application US/07876941A  
Patent No. 5885768

## GENERAL INFORMATION:

APPLICANT: Reyes, Gregory R.  
APPLICANT: Bradley, Daniel W.  
APPLICANT: Tam, Albert W.  
APPLICANT: Mitchell, Carl  
TITLE OF INVENTION: Hepatitis E Virus Peptide Antigen and  
TITLE OF INVENTION: Antibodies  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/876,941A  
FILING DATE: 01-MAY-1992  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 822,335  
FILING DATE: 17-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 505,888  
FILING DATE: 05-APRIL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 420,921  
FILING DATE: 13-OCTOBER-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 367,486  
FILING DATE: 16-JUNE-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 336,672  
FILING DATE: 11-APRIL-1989

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 208,997  
FILING DATE: 17-JUNE-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.33  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: C2, MEXICO, FIGURE 9  
US-07-876-941A-18

Query Match 80.5%; Score 33; DB 2; Length 435;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 2 RLVOPG 8  
9 RLVOPG 15

RESULT 6  
US-08-259-148A-17  
Sequence 17, Application US/08259148A  
Patent No. 574190  
GENERAL INFORMATION:  
APPLICANT: Reyes, Gregory R.  
APPLICANT: Bradley, Daniel W.  
APPLICANT: Twu, Jr-Shin  
APPLICANT: Purdy, Michael A.  
APPLICANT: Tam, Albert W.  
APPLICANT: Krawczynski, Krzysztof Z.  
APPLICANT: Yarbough, Patrice D.  
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/259,148A  
FILING DATE: 13-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 822,335  
FILING DATE: 17-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 505,888  
FILING DATE: 05-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 420,921  
FILING DATE: 13-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 367,486  
FILING DATE: 16-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 336,672  
FILING DATE: 11-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 208,997  
FILING DATE: 17-JUN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9

US-08-259-148A-17  
Query Match 80.5%; Score 33; DB 1; Length 436;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 2 RLVOPG 8  
9 RLVOPG 15

RESULT 7  
US-08-484-054-17  
Sequence 17, Application US/08484054  
Patent No. 5770689  
GENERAL INFORMATION:  
APPLICANT: Reyes, Gregory R.  
APPLICANT: Bradley, Daniel W.  
APPLICANT: Twu, Jr-Shin  
APPLICANT: Purdy, Michael A.  
APPLICANT: Tam, Albert W.  
APPLICANT: Krawczynski, Krzysztof Z.  
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,054  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 870,985  
FILING DATE: 20-APRIL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 822,335  
FILING DATE: 17-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 505,888  
FILING DATE: 05-APRIL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 420,921  
FILING DATE: 13-OCTOBER-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 367,486  
FILING DATE: 16-JUNE-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 336,672  
FILING DATE: 11-APRIL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 208,997  
FILING DATE: 17-JUNE-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.38  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9  
US-08-484-054-17

Query Match 80.5%; Score 33; DB 1; Length 436;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPE 8  
DB 9 RLVOPE 15

## RESULT 8

US-07-876-941A-17  
Sequence 17, Application US/07876941A  
Patent No. 5885768  
GENERAL INFORMATION:  
APPLICANT: Reyes, Gregory R.  
APPLICANT: Bradley, Daniel W.  
APPLICANT: Tam, Albert W.  
APPLICANT: Mitchell, Carl  
TITLE OF INVENTION: Hepatitis E Virus Peptide Antigen and  
TITLE OF INVENTION: Antibodies  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/876,941A  
FILING DATE: 01-MAY-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 822,335  
FILING DATE: 17-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 505,888  
FILING DATE: 05-APRIL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 420,921  
FILING DATE: 13-OCTOBER-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 367,486  
FILING DATE: 16-JUNE-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 336,672  
FILING DATE: 11-APRIL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 208,997  
FILING DATE: 17-JUNE-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.33  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9  
US-07-876-941A-17

Query Match 80.5%; Score 33; DB 2; Length 436;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPE 8  
DB 9 RLVOPE 15

## RESULT 9

PCT-US95-13703-27  
Sequence 27, Application PC/TUS9513703  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13703  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0293.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E virus (Burma strain)  
INDIVIDUAL ISOLATE: r62kda, 56.5 kda  
PCT-US95-13703-27

Query Match 80.5%; Score 33; DB 5; Length 525;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPE 8  
DB 122 RLVOPE 128

## RESULT 10

PCT-US95-13703-28  
Sequence 28, Application PC/TUS9513703  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13703  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0293.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E virus (Mexico strain)  
PCT-US95-13703-28

Query Match 80.5%; Score 33; DB 5; Length 525;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
1:|||||  
Db 122 RLVOPG 128

RESULT 11  
PCT-US95-13703-25  
Sequence 25, Application PC/TUS9513703  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13703  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0293.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 540 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E virus (Burma strain)  
PCT-US95-13703-25

Query Match 80.5%; Score 33; DB 5; Length 540;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
1:|||||  
Db 122 RLVOPG 128

RESULT 12  
PCT-US95-13703-26  
Sequence 26, Application PC/TUS9513703  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13703  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0293.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 540 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO

ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E virus (Mexico strain)  
INDIVIDUAL ISOLATE: r62kda, 58.1 kda  
PCT-US95-13703-26

Query Match 80.5%: Score 33; DB 5; Length 549;  
Best Local Similarity 85.7%: Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
1:|||||  
Db 122 RLVOPG 128

RESULT 13  
PCT-US95-13703-15  
Sequence 15, Application PC/TUS9513703  
GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND  
TITLE OF INVENTION: USES THEREFOR  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13703  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0293.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 549 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E virus (Burma strain)  
INDIVIDUAL ISOLATE: r62kda, FIGURE 4  
PCT-US95-13703-15

Query Match 80.5%: Score 33; DB 5; Length 549;  
Best Local Similarity 85.7%: Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
1:|||||  
Db 122 RLVOPG 128

RESULT 14  
PCT-US95-13703-16  
Sequence 16, Application PC/TUS9513703  
GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND  
TITLE OF INVENTION: USES THEREFOR  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13703  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0293.41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 549 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E virus (Mexico strain)  
INDIVIDUAL ISOLATE: r62kda, FIGURE 4  
PCT-US95-13703-16

Query Match 80.5%: Score 33; DB 5; Length 549;  
Best Local Similarity 85.7%: Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RLVOPG 8  
1:|||||  
Db 122 RLVOPG 128

RESULT 15  
US-08-240-049B-16  
Sequence 16, Application US/08240049B  
Patent No. 5686239  
GENERAL INFORMATION:  
APPLICANT: Reyes, Gregory R.  
APPLICANT: Tam, Albert W.  
TITLE OF INVENTION: Hepatitis E Virus Peptides and Methods  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penndscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/240,049B  
FILING DATE: 09-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Charles K. Sholtz  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0093.35; G32P4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hepatitis E Virus (Mexico Strain)  
INDIVIDUAL ISOLATE: ORF-2  
US-08-240-049B-16

Query Match 80.5%; Score 33; DB 1; Length 659;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLWOPG 8  
|:|||||  
Db 233 RLWOPG 239

Search completed: June 13, 2001, 14:16:33  
Job time: 494 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:33 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-8

Perfect score: 40

Sequence: 1 GRLLVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: A\_Geneseq\_0401.\*

1: /SID6/gcgdata/geneseq/AA1980.DAT.\*  
2: /SID6/gcgdata/geneseq/AA1981.DAT.\*  
3: /SID6/gcgdata/geneseq/AA1982.DAT.\*  
4: /SID6/gcgdata/geneseq/AA1983.DAT.\*  
5: /SID6/gcgdata/geneseq/AA1984.DAT.\*  
6: /SID6/gcgdata/geneseq/AA1985.DAT.\*  
7: /SID6/gcgdata/geneseq/AA1986.DAT.\*  
8: /SID6/gcgdata/geneseq/AA1987.DAT.\*  
9: /SID6/gcgdata/geneseq/AA1988.DAT.\*  
10: /SID6/gcgdata/geneseq/AA1989.DAT.\*  
11: /SID6/gcgdata/geneseq/AA1990.DAT.\*  
12: /SID6/gcgdata/geneseq/AA1991.DAT.\*  
13: /SID6/gcgdata/geneseq/AA1992.DAT.\*  
14: /SID6/gcgdata/geneseq/AA1993.DAT.\*  
15: /SID6/gcgdata/geneseq/AA1994.DAT.\*  
16: /SID6/gcgdata/geneseq/AA1995.DAT.\*  
17: /SID6/gcgdata/geneseq/AA1996.DAT.\*  
18: /SID6/gcgdata/geneseq/AA1997.DAT.\*  
19: /SID6/gcgdata/geneseq/AA1998.DAT.\*  
20: /SID6/gcgdata/geneseq/AA1999.DAT.\*  
21: /SID6/gcgdata/geneseq/AA2000.DAT.\*  
22: /SID6/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	8	21	Peptide antagonist
2	37	92.5	8	21	Peptide antagonist
3	35	87.5	8	21	Peptide antagonist
4	35	87.5	399	13	Zonula occludens t
5	33	82.5	8	21	Peptide antagonist
6	33	82.5	8	21	Peptide antagonist
7	32	80.0	8	21	Peptide antagonist
8	32	80.0	8	21	Peptide antagonist
9	32	80.0	2391	15	Carbamoyl-phosphat
10	31	77.5	203	16	Carnitine racemase
11	31	77.5	203	16	E.coli L-carnitine

12	31	77.5	390	21	B41640
13	31	77.5	415	21	G59077
14	31	77.5	415	21	G59619
15	31	77.5	442	21	G23560
16	31	77.5	449	21	G59076
17	31	77.5	463	21	G59618
18	31	77.5	493	21	G33559
19	31	77.5	495	21	G23558
20	31	77.5	495	21	G51350
21	31	77.5	528	21	G51349
22	31	77.5	548	21	G51348
23	31	77.5	575	21	G59617
24	31	77.5	590	21	G59075
25	30	75.0	8	21	G79107
26	30	75.0	8	21	G79120
27	30	75.0	117	16	R66214
28	30	75.0	117	17	R81470
29	30	75.0	137	21	G05180
30	30	75.0	144	22	B19779
31	30	75.0	149	21	G36079
32	30	75.0	149	21	G44702
33	30	75.0	151	21	G23109
34	30	75.0	165	21	G19553
35	30	75.0	166	21	B28135
36	30	75.0	166	21	G07728
37	30	75.0	166	21	G11107
38	30	75.0	166	21	G20486
39	30	75.0	166	21	G23108
40	30	75.0	166	21	G45133
41	30	75.0	166	21	G45692
42	30	75.0	166	21	G45717
43	30	75.0	168	21	G36078
44	30	75.0	168	21	G44701
45	30	75.0	180	21	G45132

## ALIGNMENTS

RESULT 1	
ID Y79112	Y79112 standard; Peptide; 8 AA.
XX	
AC Y79112;	
XX	
DT 05-JUN-2000	(first entry)
XX	
DE	Peptide antagonist of zonulin.
KW	Zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antidiarr; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnary; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
OS	Synthetic.
XX	
PN WO200007609-A1.	
XX	
PD 17-FEB-2000.	
XX	
PF 28-JUL-1999;	99WC-US16683.
XX	
PR 03-AUG-1998;	98US-0127815.
XX	
PA (UYMA-) UNIV MARYLAND BALTIMORE.	
XX	
PI Fasano A.	
XX	
DR WPI; 2000-205565/18.	
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1; Page 43; 69pp; English.

Claim 1; Page 43; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC anti-inflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

**SQ Sequence 8 AA;**

Query Match	100.0%;	Score 40;	DB 21;	Length 8;
Best Local Similarity	100.0%;	Pred. No. 3.2e+05;		
Matches	8;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0

QY	1	GRLVQDG	8
Db	1	grllvqdg	8

RESULT	2
Y79108	
ID	Y79108 standard; Peptide; 8 AA.

DT 05-JUN-2000 (first entry)

Peptide antagonist of zonulin.

KM zonulin; antagonist; zonula occludens toxin receptor;  
KM neuroprotective; antiinflammatory; cerebroprotective;  
KM blood-brain barrier; antitumor; antidiabetic; antiviral;  
KM neuroprotective; dermatological; antitumor; antiviral;  
KM antibacterial; cytostatic; anti-HIV; vulnereary; antiallergic;  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KM gastrointestinal inflammation; therapy.

OS Synthetic

PN WO200007609-A1.

PD 17-FEB-2000

PF 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.

PA (UYMA-) UNIV MARYLAND BALTIMORE.

PI	Fasano A;
XX	
DR	WPI; 2000-205565/18.
XX	

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
XX  
PS Claim 1; Page 42; 69pp; English.

Claim 1; Page 42; 69pp; English.

This present sequence is that of a peptide antagonist of zonulin (2), one of 25 such peptides (see YV9105-29) of the invention, which bind to a zonula occludens toxin (ZOT) receptor, yet do not physiologically modulate the opening of mammalian tight junctions (TJ). The peptide antagonists are based on a common motif of ZOT and human zonulins, which is believed to be critical for receptor binding. They can be prepared by chemical synthesis or by use of recombinant DNA techniques. The peptide antagonists are used as antiinflammatory agents in the treatment of gastrointestinal inflammation, where they bind to the ZOT receptor in the intestine and yet does not physiologically modulate the opening of TJ in the intestine. Gastrointestinal inflammation conditions give rise to increased intestinal permeability and the peptide is useful for treating intestinal conditions that cause protein losing enteropathy caused by infection, e.g. Clostridium difficile infection, enterocolitis, shigellosis, viral gastroenteritis, parasite infestation, bacterial overgrowth, whipple's disease, diseases with mucosal erosion or ulcerations, e.g. gastritis, gastric cancer, collagenous colitis, inflammatory bowel disease, diseases marked by lymphatic obstruction, e.g. congenital intestinal lymphangiectasia, sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical correction of congenital heart disease with Fontan's operation, mucosal diseases without ulceration, e.g. Menetrier's disease, coeliac disease, eosinophilic gastroenteritis, and immune diseases, e.g. systemic lupus erythematosus or food allergies, primarily to milk.

Sequence 8 AA;

Query Match	92.5%	Score 37	DB 21	Length 8
Best Local Similarity	87.5%	Pred. No. 3.2e+05		
Matches	7	Conservative	1	Mismatches 0; Indels 0; Gaps 0

Qy	1	GRLVQDG	8
		:	
Db	1	grv\vgdq	8

RESULT	3
Y79110	
ID	Y79110 standard; Peptide; 8 AA.

DT 05-JUN-2000 (first entry)  
 XX

Peptide antagonist of zonulin.

KM zonulin; antagonist; zonula occludens toxin receptor;  
KM blood-brain barrier; antiinflammatory; cerebroprotective;  
KM neuroprotective; dermatological; antileuc; antiviral;  
KM antibacterial; cytosstatic; anti-HIV; vulnerary; antiallergic;  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KM gastrointestinal inflammation; therapy.

05 Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.



XX PR 03-AUG-1998; 98US-0127815.  
 XX PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX PI Fasano A;  
 XX DR WPI: 2000-205565/18.  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
 XX  
 PS Claim 1; Page 42; 69pp; English.  
 XX  
 CC This present sequence is that of a peptide antagonist of zonulin  
 CC (4), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 XX  
 SQ Sequence 8 AA;  
 XX  
 Query Match 87.5%; Score 35; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. NO. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRLLVODG 8  
 DB 1 grlcvcqdg 8  
 XX  
 RESULT 4  
 R20006  
 ID R20006 standard; Protein; 399 AA.  
 XX  
 AC R20006;  
 XX  
 DT 31-MAR-1992 (first entry)  
 XX  
 DE Zonula occludens toxin.  
 XX  
 KW ZOT; cholera; vaccine; enterotoxin; diarrhoea.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN MO9118979-A.  
 XX  
 PD 12-DEC-1991.  
 XX  
 PF 05-JUN-1991; 91WO-US03812.

XX PR 05-JUN-1990; 90US-0533315.  
 XX PA (UYMA-) UNIV MARYLAND BALTI.  
 XX PI Kaper JB, Baudry-Maurelli B, Fasano A;  
 XX DR WPI: 1992-007465/01.  
 XX DR N-PSDB; Q20185.  
 XX  
 PT New Vibrio cholerae strains - comprise restriction endonuclease  
 PT fragment encoding toxin, used as vaccines against cholera  
 XX  
 PS Disclosure; Fig 18; 83pp; English.  
 XX  
 CC The amino acid sequence is that of the zonula occludens toxin (ZOT).  
 CC It may be responsible for diarrhoea in some strains of cholera and  
 CC the ZOT gene or fragments of it are deleted from strains of Vibrio  
 CC cholerae (V.c.) to be used as vaccines. These V.c. strains have 100%  
 CC efficacy in protecting humans against subsequent infection with a  
 CC strain of a similar serotype and avoid undesirable side effects such  
 CC as diarrhoea, nausea and cramping. Cultures of these strains may be  
 CC used for prodn. of vaccines against cholera.  
 XX  
 SQ Sequence 399 AA;  
 XX  
 Query Match 87.5%; Score 35; DB 13; Length 399;  
 Best Local Similarity 87.5%; Pred. NO. 30;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRLLVODG 8  
 DB 291 grlcvcqdg 298  
 XX  
 RESULT 5  
 Y79111  
 ID Y79111 standard; Peptide; 8 AA.  
 XX  
 AC Y79111;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Peptide antagonist of zonulin.  
 XX  
 KW Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnereary; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200007609-A1.  
 XX  
 PD 17-FEB-2000.  
 XX  
 PF 28-JUL-1999; 99WO-US16683.  
 XX  
 PR 03-AUG-1998; 98US-0127815.  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX  
 PI Fasano A;  
 XX  
 DR WPI: 2000-205565/18.  
 XX  
 PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
 XX

PS Claim 1; Page 43; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulin, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA:

Query Match 82.5%; Score 33; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVODG 8  
 | | | | | | |  
 Db 1 grllvpg 8.

RESULT 6  
 ID Y79124 standard; Peptide: 8 AA.

AC Y79124;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 XX blood-brain barrier; antiinflammatory; cerebroprotective;  
 XX neuroprotective; dermatological; antilucer; antiviral;  
 XX antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
 XX hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 XX gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.

PA (UYMA-) UNIV MARYLAND BALTIMORE.

PI Fasano A;

PR WPI; 2000-205565/18.

XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 PS Claim 1; Page 47; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulin, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.

SO Sequence 8 AA:

Query Match 82.5%; Score 33; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVODG 8  
 | | | | | | |  
 Db 1 g9llvqdg 8

RESULT 7  
 ID Y79106 standard; Peptide: 8 AA.

AC Y79106;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 XX blood-brain barrier; antiinflammatory; cerebroprotective;  
 XX neuroprotective; dermatological; antilucer; antiviral;  
 XX antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
 XX hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 XX gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.



```

FT FT Domain /note= "glutamine-amidotransferase domain"
FT FT 1..270
FT FT /note= "structural subdomain"
FT FT Peptide 271..482
FT FT /note= "insert sequence"
FT FT Domain 483..690
FT FT /note= "glutaminase subdomain"
FT FT Domain 691..2391
FT FT /note= "carbamoyl-phosphate-synthase domain"
FT FT Peptide 691..1254
FT FT /note= "ATP binding subdomain CPsa"
FT FT Peptide 1255..1857
FT FT /note= "insert sequence"
FT FT Domain 1858..2391
FT FT /note= "ATP binding subdomain CPSP"
XX PN W09412643-A.
XX PD 09-JUN-1994.
XX PE 02-DEC-1993; 93WO-AU00617.
XX PF 03-DEC-1992; 92AU-0006206.
XX PR 16-DEC-1992; 92AU-0006380.
XX PA (UNIX ) UNISEARCH LTD.
XX PI Flores MV, Osullivan WJ, Stewart TS;
XX DR WPI: 1994-200271/24.
XX DR N-PSDB: 062924.
XX PT Nucleic acid encoding carbamoyl phosphate synthetase II -
XX PT isolated from Plasmodium falciparum, used to develop prods. for
XX PT the treatment of malaria.
XX PS Disclosure; Page 6-16; 31pp; English.
XX CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II
XX CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes
XX CC a protein that includes 2 insert sequences not found in other CPSII
XX CC proteins. The first separates the putative structural subdomain and
XX CC the glutaminase subdomain of the glutamine-amidotransferase subunit
XX CC of CPSII, while the second separates 2 ATP binding subdomains of the
XX CC CPSII subunit, CPsa and CPSP.
XX CC CSII subunit, CPsa and CPSP.
XX SQ Sequence 2391 AA:

Query Match 80.0%; Score 32; DB 15; Length 2391;
Best Local Similarity 62.5%; Pred. No. 7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVODG 8
DB 29 grlledg 36

RESULT 10
R80284
ID R80284 standard; Protein; 203 AA.
XX AC R80284;
XX DT 08-MAR-1996 (first entry)
XX DE Carnitine racemase co-factor.
XX KM caid gene; cai operon; carnitine racemase; L-carnitine; crotonobetaine;
XX KM co-factor; cai; transport; long chain activated fatty acid; treatment;
XX KM deficiency state; muscular dystrophy; lipid storage myopathy.
XX OS Escherichia coli, strain 044K74.

```

```

XX PN DE4402127-A1.
XX PD 27-JUL-1995.
XX PE 21-JAN-1994; 94DE-4402127.
XX PR 21-JAN-1994; 94DE-4402127.
XX PA (UYLE ) UNIV LEIPZIG.
XX PI Eichler K, Kleber H, Mandrand M;
XX DR WPI: 1995-264414/35.
XX DR N-PSDB: Q98566.
XX PT New carnitine racemase from E.coli and related plasmids - for prodn.
XX PT of L-carnitine from its D-isomer, e.g. for treatment of carnitine
XX PT deficiency, stimulating prodn. of monoclonal antibodies, etc.
XX PS Claim 2; Page 9; 10pp; German.
XX CC Carnitine racemase (CR) co-factor is encoded by Q98566, the open reading
XX CC frame from the cai gene of the cai operon of E. coli 044K74 (DSM 8828)
XX CC CR is encoded by Q98565 (the open reading frame from the cai gene of the
XX CC same operon). The cai gene is induced by L-carnitine or crotonobetaine
XX CC (but not by D-carnitine) under aerobic conditions and requires the cai
XX CC co-factor for activity. CR is used to prepare L-carnitine from its
XX CC D-isomer. L-carnitine is involved in transport of long chain activated
XX CC fatty acids across internal mitochondrial membranes and is useful for
XX CC treating deficiency states, e.g. muscular dystrophy or lipid storage
XX CC myopathy, as well as chronic kidney insufficiency.
XX SQ Sequence 203 AA:

Query Match 77.5%; Score 31; DB 16; Length 203;
Best Local Similarity 75.0%; Pred. No. 95;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVODG 8
DB 55 grllvqg 62

RESULT 11
R80278
ID R80278 standard; Protein; 203 AA.
XX AC R80278;
XX DT 28-FEB-1996 (first entry)
XX DE E.coli L-carnitine dehydratase accessory protein caiE.
XX KM E.coli; L-carnitine dehydratase; caiB; carnitine operon; CaiE; rat;
XX KM intestine; chromosomal DNA library; probe; expression plasmid; Bacterium;
XX KM crotonobetaine; long chain fatty acid; transport; mitochondrial membrane;
XX KM haemodialysis; cardiac.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1
XX FT /note= "encoded by crc"
XX PN FR2715167-A1.
XX PD 21-JUL-1995.
XX PF 20-JAN-1994; 94FR-0000807.
XX PR 20-JAN-1994; 94FR-0000807.

```

XX (MASC-) INST NAT SCI APPLIQUEES LYON.  
 PA  
 XX  
 PI Elchler K, Kieber H, Mandrand-Berthelot M;  
 XX  
 DR WPI: 1995-256580/34.  
 XX  
 DR N-PSDB: Q98554.  
 XX  
 PT DNA fragments encoding E.coli L-carnitine dehydratase and Caie  
 PT protein - and related plasmids, transformed E.coli and recombinant  
 PT proteins, useful in enantioselective prodn of L-carnitine  
 XX  
 PS Claim 10; Page 22; 32pp; French.  
 XX  
 CC The amino acid sequence of the E.coli L-carnitine dehydratase (CDH)  
 CC accessory protein Caie encoded by the gene designated caie. The sequence  
 CC forms part of the carnitine operon which includes the L-carnitine  
 CC dehydratase gene caib (Q98553). The sequences were obtained from  
 CC E.coli strain 044K74 which was isolated from rat intestine. The caib  
 CC gene was obtained by screening an E.coli chromosomal DNA library with the  
 CC probes Q98555-7, based on the N-terminal sequence derived from the  
 CC purified enzyme. The 1.3 kb MluI-MluII caib gene fragment was subcloned  
 CC into the expression plasmid pT7-6 for production of the enzyme in E.coli.  
 CC Bacteria expressing the L-carnitine dehydratase can be used in the  
 CC microbiological prodn. of L-carnitine from crotonobetaine. L-carnitine  
 CC is an essential factor in long chain fatty acid transport across internal  
 CC mitochondrial membranes. It can be used to treat L-carnitine-deficiency  
 CC syndromes, haemodialysis and cardiac patients and to stimulate prodn. of  
 CC monoclonal antibodies.  
 XX  
 SQ Sequence 203 AA;  
 OY  
 DB 1 GRLVODG 8  
 1111111  
 55 grlvvgtg 62  
 RESULT 12  
 B41640  
 ID B41640 standard; Protein: 390 AA.  
 AC B41640;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1404 polypeptide sequence SEQ ID NO:2808.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
 KW vulnerability; antiparotatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiac;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antineuritic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach M;  
 XX  
 DR WPI: 2000-602362/57.  
 XX  
 DR N-PSDB: C75849.  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 2046-2047; 5507pp; English.  
 XX  
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerability;  
 CC antiparotatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;  
 CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antineoplastic; antithyroid;  
 CC antiviral; antifungal; antirheumatic; antineuritic; antianemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 390 AA;  
 OY  
 DB 2 RLIVODG 8  
 1111111  
 247 rllvedg 253  
 RESULT 13  
 G59077  
 ID G59077 standard; Protein: 415 AA.  
 AC G59077;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 76377.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125708.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135214.  
PR 21-MAY-1999; 99US-0135333.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0146381.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158363.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 415;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GRLLVQDG 8  
||:|:|:|  
Db 218 grllrdg 225

RESULT 14  
G59619  
ID G59619 standard; Protein: 415 AA.

XX G59619;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 77133.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147312.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 415;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Caps 0;

QY 1 GRLLVODG 8  
11:1:11  
Db 218 grillrdg 225

RESULT 15  
ID G23560  
AC G23560;  
XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 26913.  
XX  
XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

XX  
XX  
PN EP1033405-A2.

PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.



PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0133247.  
PR 04-MAY-1999; 99US-0133484.  
PR 05-MAY-1999; 99US-0133485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0138459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 77.58; Score 31; DB 21; Length 442;  
Best Local Similarity 85.7%; Pred. NO. 2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Caps 0;

OY 2 RLIVODG 8  
1:|||||  
Db 250 rvlvqdg 256

Search completed: June 13, 2001, 14:14:34  
Job time: 375 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:41 ; Search time 130.61 Seconds  
(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825a-8

Perfect score: 40

Sequence: 1 GRLLVQDG 8

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	87.5	399	2	B82197
2	35	87.5	399	2	A43864
3	35	87.5	687	2	T29148
4	34	85.0	1317	2	S77517
5	33	82.5	180	2	S09888
6	33	82.5	382	2	H84025
7	33	82.5	482	2	H69392
8	33	82.5	495	2	S76967
9	32	80.0	363	2	A83177
10	32	80.0	640	2	T34916
11	32	80.0	652	2	B82724
12	32	80.0	893	2	T28621
13	32	80.0	1018	1	S73720
14	32	80.0	1920	2	A53188
15	32	80.0	2391	2	T18410
16	31	77.5	196	2	C64891
17	31	77.5	228	2	T50590
18	31	77.5	390	2	B55889
19	31	77.5	427	2	G69067
20	31	77.5	467	2	S15297
21	31	77.5	590	2	T06626
22	31	77.5	4063	2	T42993
23	30	75.0	4101	2	T23630
24	30	75.0	160	2	A75466
25	30	75.0	161	2	S40872
26	30	75.0	182	2	A64834
27	30	75.0	203	2	C64724
28	30	75.0	225	1	E53402
29	30	75.0	239	2	A75164

30	30	75.0	280	2	F70305	spermidine synthase
31	30	75.0	288	2	S40173	ExoC protein - Rhl
32	30	75.0	301	2	G70426	aminoethyltransferase
33	30	75.0	318	2	T07095	2'-hydroxyisotriazole
34	30	75.0	318	2	S17744	2'-hydroxyisotriazole
35	30	75.0	318	2	S48631	2'-hydroxyisotriazole
36	30	75.0	322	2	H64933	hypothetical prote
37	30	75.0	342	2	D49348	succinoglycan bios
38	30	75.0	347	2	A47370	GTP-binding regula
39	30	75.0	355	2	A26984	std protein - Scr
40	30	75.0	365	2	H69231	sensory transducti
41	30	75.0	402	2	T09062	probable advanced
42	30	75.0	433	2	A70536	probable pepc prot
43	30	75.0	434	2	D83456	probable guanine d
44	30	75.0	441	2	G71508	glucose-1-phosphat
45	30	75.0	441	2	F81667	glucose-1-phosphat

#### ALIGNMENTS

```
RESULT 1
B82197
zona occludens toxin VC1458 [imported] - Vibrio cholerae (group O1 strain N16961)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: B82197
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MID:20406833
A:Accession: B82197
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <HEI>
A:Cross-references: GB:AE004224; GB:AE003852; NID:g9655952; PIDN:AAF94615.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1458
A:Map position: 1

Query Match 87.5%; Score 35; DB 2; Length 399;
Best Local Similarity 87.5%; Pred. No. 7.7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVQDG 8
DB 291 GRLLVQDG 298

RESULT 2
A43864
zona occludens toxin - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A43864
R:Baudry, B.; Pasano, A.; Kettley, J.; Kaper, J.B.
Infec. Immun. 60, 428-434, 1992
A:Title: Cloning of a gene (zot) encoding a new toxin produced by Vibrio cholerae.
A:Reference number: A43864; MID:92112300
A:Accession: A43864
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <BAU>
A:Cross-references: GB:M83563; NID:g155314; PIDN:AAA27582.1; PID:g155315
A:Note: sequence extracted from NCBI backbone (NCBIN:77488, NCBIPI:77491)

Query Match 87.5%; Score 35; DB 2; Length 399;
Best Local Similarity 87.5%; Pred. No. 7.7;
```

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVODG 8  
||| ||||  
Db 291 GRLCVODG 298

RESULT 3  
T29148  
hypothetical protein F56A3.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T29148  
R:Gelsel, C.; Bradshaw, H.; Wamsley, P.  
submitted to the EMBL Data Library, November 1996  
A:Description: The sequence of C. elegans cosmid F56A3.  
A:Reference number: 220578  
A:Accession: T29148  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-687 <GB1>  
A:Cross-references: EMBL:U80446; PIDN:AB37805.1; GSPDB:GN00019; CESP:F56A3.1  
A:Experimental source: strain Bristol N2; clone F56A3  
C:Genetics:  
A:Gene: CESP:F56A3.1  
A:Map position: 1  
A:Insertions: 103/2; 132/3; 169/2; 269/3; 317/2; 464/1; 522/3; 582/3; 612/2; 654/2; 687/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein F56A3.1

Query Match 87.5%; Score 35; DB 2; Length 687;  
Best Local Similarity 87.5%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVODG 8  
||| ||||  
Db 587 GRLVODG 594

RESULT 4  
S77517  
DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain - Synechocystis sp. (strain PCC 7680)  
N:Alternate names: protein sll1789  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S77517  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
5.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S77517  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1317 <KAN>  
A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BAAL7364.1; PID:g165244  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: rpoC2  
C:Superfamily: chloroplast DNA-directed RNA polymerase beta'-2 chain  
C:Keywords: nucleotidyltransferase; transcription

Query Match 85.0%; Score 34; DB 2; Length 1317;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLIVODG 8  
|||||||  
Db 840 RLIVODG 846

RESULT 5  
S09888  
hypothetical protein UL121 precursor - human cytomegalovirus (strain AD169)

C:Species: human cytomegalovirus, human herpesvirus 5  
A:Note: host Homo sapiens (man)  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999  
C:Accession: S09888  
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerry, R.; Horsnell, T.  
M.; Barrrell, B.G.  
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
A:Reference number: S09749; MUID:90265039  
A:Accession: S09888  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-180 <CHE>  
A:Cross-references: EMBL:X17403; NID:g59591; PIDN:CA35323.1; PID:g1780903  
A:Note: this sequence was submitted to the EMBL Data Library, December 1989  
C:Superfamily: human cytomegalovirus hypothetical protein UL121  
C:Keywords: glycoprotein; transmembrane protein  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:36-180/Product: hypothetical protein UL121 #status predicted <MAT>  
F:142-164/Domain: transmembrane #status predicted <TM>  
F:104/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 82.5%; Score 33; DB 2; Length 180;  
Best Local Similarity 85.7%; Pred. No. 9;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLIVODG 8  
|||||||  
Db 108 RLIVODG 114

RESULT 6  
H84025  
hypothetical protein BH308 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: H84025  
R:Takani, H.; Nakasone, K.; Takai, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: AB3650; MUID:20263314  
A:Accession: H84025  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1382 <STO>  
A:Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BA06727.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH3008

Query Match 82.5%; Score 33; DB 2; Length 382;  
Best Local Similarity 75.0%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVODG 8  
||| ||||  
Db 20 GRLVODG 27

RESULT 7  
H69392  
4-hydroxybutyrate CoA transferase (cat2-1) homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-2000  
C:Accession: H69392  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weldman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Arlisch, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: H69392  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-482 <KLE>  
 A:Cross-references: GB:AE001025; GB:AE000782; NID:g2689348; PIDN:AAB90101.1; PID:g264944  
 C:Superfamily: acetyl-CoA hydrolase

Query Match 82.5%; Score 33; DB 2; Length 482;  
 Best Local Similarity 75.0%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVODG 8  
 |||:||||  
 Db 225 GRLLVODG 232

RESULT 8  
 S76967  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C:Accession: S76967  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 S.

A:Reference number: S74322; MUID:97061201  
 A:Accession: S76967  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-495 <KAN>  
 A:Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BA10659.1; PID:d101131  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 82.5%; Score 33; DB 2; Length 495;  
 Best Local Similarity 85.7%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOD 7  
 ||||:||  
 Db 451 GRLLVOD 457

RESULT 9  
 AB3177  
 Probable N-acetylglucosamine-6-phosphate deacetylase PA3758 [Imported] - Pseudomonas aer  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: AB3177  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 edman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Llm,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437337  
 A:Accession: AB3177  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-363 <STO>  
 A:Cross-references: GB:AE004794; GB:AE004091; NID:g9949917; PIDN:AA07145.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:

Query Match 80.0%; Score 32; DB 2; Length 363;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVODG 8  
 |||:||||  
 Db 15 GRLLVODG 22

RESULT 10  
 T34916  
 transferase - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 02-Sep-2000  
 C:Accession: T34916  
 R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, January 1998  
 A:Reference number: Z21558  
 A:Accession: T34916  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-640 <COLI>  
 A:Cross-references: EMBL:AL021409; PIDN:CA16181.1; GSPDB:GN00070; SCOEDB:SC3F7.10  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SC3F7.10  
 C:Superfamily: glycine C-acetyltransferase homology  
 F:287-624/Domain: glycine C-acetyltransferase homology <GCA>

Query Match 80.0%; Score 32; DB 2; Length 640;  
 Best Local Similarity 75.0%; Pred. No. 58;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVODG 8  
 ||||:||  
 Db 415 GRLLVODG 422

RESULT 11  
 B82724  
 cardiolipin synthase xrl087 [Imported] - Xylella fastidiosa (strain 965c)  
 C:Species: Xylella fastidiosa.  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: B82724  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: AB2515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82724  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-652 <SMT>  
 A:Cross-references: GB:AE003945; GB:AE003849; NID:g9106036; PIDN:AAF83897.1; GSPDB:GN  
 A:Experimental source: strain 965c  
 R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.V.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
 as-Neto, E.; Docena, C.; El-Doriry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to Genbank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
 J.D.; Dungeleira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
 A:Authors: Martins, E.M.F.; Matsukuma, A.V.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
 .; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
 M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:

A:Gene: XF1087

Query Match 80.0%; Score 32; DB 2; Length 652;  
Best Local Similarity 62.5%; Pred. No. 59;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVQDG 8  
|||||  
Db 260 GRMVLQDG 267

RESULT 12

T29621  
hypothetical protein T05E8.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T29621

R:Wu, X.; Le, T.T.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of *C. elegans* cosmid T05E8.

A:Reference number: 220654

A:Accession: T29621

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-893 <WUX>

A:Cross-references: EMBL:U97014; PIDN:AB52425.1; GSPDB:GN00019; CESP:T05E8.1

A:Experimental source: strain Bristol N2; clone T05E8

C:Genetics:

A:Gene: CESP:T05E8.1

A:Map position: 1

A:Introns: 22/3; 87/3; 185/3; 244/3; 297/3; 367/2; 404/3; 545/2; 673/3

Query Match 80.0%; Score 32; DB 2; Length 893;  
Best Local Similarity 62.5%; Pred. No. 83;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVQDG 8  
|||||  
Db 78 GRIVRDG 85

RESULT 13

S73720  
cytochrome accessory protein HMW1 - *Mycoplasma pneumoniae* (strain ATCC 29342)

N:Alternate names: hypothetical protein H08\_orf1018

C:Species: *Mycoplasma pneumoniae*

A:Variety: ATCC 29342

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 10-Dec-1999

C:Accession: S73720; S49064; S49065

R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkli, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A>Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*

A:Reference number: S73327; MUID:97105885

A:Accession: S73720

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1018 <HIM>

A:Cross-references: EMBL:AF000038; GB:U00089; NID:91674074; PIDN:AB96042.1; PID:9167407

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

R:Prof. T.; Herrmann, R.

Mol. Microbiol. 13, 337-348, 1994

A>Title: Identification and characterization of hitherto unknown *Mycoplasma pneumoniae* P

A:Reference number: S49059; MUID:95075318

A:Accession: S49064

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 127-242 <PRO>

A:Cross-references: EMBL:232661; NID:9474075; PIDN:CAA83580.1; PID:9474076

A:Experimental source: clone H8-1P

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994

A:Accession: S49065

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 62-113 <PRM>

A:Cross-references: EMBL:232662; NID:9474077; PIDN:CAA83581.1; PID:9474078

A:Experimental source: clone E9-4B

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994

C:Genetics:

A:Gene: hmw1

A:Genetic code: SGC3

C:Superfamily: cytochrome-accessory protein hmw1

Query Match 80.0%; Score 32; DB 1; Length 1018;  
Best Local Similarity 75.0%; Pred. No. 96;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVQDG 8  
|||||  
Db 29 GRLTVQDG 36

RESULT 14

A53188

pericentrin - mouse

C:Species: *Mus musculus* (house mouse)

C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 05-Nov-1999

C:Accession: A53188

R:Doxsey, S.J.; Stein, P.; Evans, L.; Calarco, P.D.; Kirschner, M.

Cell 76, 639-650, 1994

A>Title: Pericentrin, a highly conserved centrosome protein involved in microtubule o

A:Reference number: A53188; MUID:94170365

A:Accession: A53188

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1920 <DOX>

A:Cross-references: GB:U05023; NID:9458667; PIDN:AA17886.1; PID:9458668

C:Keywords: coiled coil

Query Match 80.0%; Score 32; DB 2; Length 1920;  
Best Local Similarity 75.0%; Pred. No. 1,9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLVQDG 8  
|||||  
Db 818 GHLVQDG 825

RESULT 15

T18410

carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) II - *malaria* parasit

C:Species: *Plasmodium falciparum*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T18410

R:Flores, M.V.C.; O'Sullivan, W.J.; Stewart, T.S.

submitted to the EMBL Data Library, March 1997

A:Description: Characterisation of the carbamoyl phosphate synthetase gene from *Plasm*

A:Reference number: Z18931

A:Accession: T18410

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2391 <FLD>

A:Cross-references: EMBL:U32150; NID:9476023; PID:9476024; PIDN:AAA29522.1

C:Superfamily: rudimentary enzyme: aspartate/ornithine carbamoyltransferase homology;

arabamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-P

C:Keywords: ligase

F;96-2367/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>

Query Match 80.0%; Score 32; DB 2; Length 2391;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Wed Jun 13 15:01:22 2001

pct-us01-05825a-8.rpr

Page 5

OY 1 GRLLVQDG 8  
| | | | |  
Db 29 GRLLVQDG 36

Search completed: June 13, 2001, 14:10:41  
Job time: 142 sec

*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:42 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-8  
Perfect score: 40  
Sequence: 1 GRLVQDC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	87.5	399	1 ZOT_VIRCH	P38442 vibrio chol
2	34	85.0	586	1 Y4GM_RHISN	P55449 rhizobium s
3	34	85.0	1317	1 RPOD_SYNY3	P73334 synechocyst
4	33	82.5	180	1 ULC1_HCMVA	P16741 human cytom
5	32	80.0	1018	1 HMA1_MYCPN	O50365 mycoplasma
6	32	80.0	1920	1 PCNT_MOUSE	P48725 mus musculu
7	31	77.5	101	1 DMT1_ALIMI	O9pue0 alligator m
8	31	77.5	196	1 PAAY_ECOLI	P7181 escherichia
9	31	77.5	205	1 DMT1_TRASC	P57650 trachemys s
10	31	77.5	311	1 DMT1_CHICK	O9pue7 gallus gall
11	31	77.5	390	1 IADL_ECOLI	P39377 escherichia
12	31	77.5	427	1 YF05_METH	O27349 methanobact
13	31	77.5	467	1 WCMG_SALTY	P26389 salmonella
14	30	75.0	161	1 MENG_ECOLI	P32185 escherichia
15	30	75.0	182	1 YCBO_ECOLI	P75855 escherichia
16	30	75.0	196	1 CAIE_ECOLI	P39206 escherichia
17	30	75.0	280	1 SPEE_AQUAE	O66473 aquifex aeo
18	30	75.0	318	1 IFR_MEDSA	P52575 medicago sa
19	30	75.0	318	1 IFR_PEA	P52576 pisum sativ
20	30	75.0	322	1 ASTE_ECOLI	P33700 rhizobium m
21	30	75.0	342	1 EXOU_RHIME	P33700 rhizobium m
22	30	75.0	347	1 GBB_DICDI	P36408 dictyostell
23	30	75.0	350	1 GCST_AQUAE	O67441 aquifex aeo
24	30	75.0	355	1 STRD_STRGR	P08070 streptomyce
25	30	75.0	403	1 RAGE_MOUSE	O62131 mus musculu
26	30	75.0	411	1 IHH_HUMAN	O14623 homo sapien
27	30	75.0	464	1 WCMG_ECOLI	P71244 escherichia
28	30	75.0	583	1 NDC1_RABIT	O28615 cryctolagus
29	30	75.0	883	1 HSS2_MOUSE	P52850 mus musculu
30	30	75.0	1052	1 POLG_TEV	P04517 t genome po
31	29	72.5	162	1 MENG_HAEIN	P44738 haemophilus
32	29	72.5	176	1 FRIM_SALISA	P49947 salmo salar
33	29	72.5	217	1 CYSE_BACSU	O06750 bacillus su

## ALIGNMENTS

RESULT 1	ID	ZOT_VIRCH	STANDARD:	PRT:	399 AA.
AC	P38442	Q9L706; Q9R3V6;			
DT	01-OCF-1994	(Rel. 30, Created)			
DT	01-OCF-2000	(Rel. 40, Last sequence update)			
DT	01-OCF-2000	(Rel. 40, Last annotation update)			
DE	ZONA OCCUDENS TOXIN (ZONULAR OCCUDENS TOXIN).				
GN	ZOT OR VC1458.				
OS	Vibrio cholerae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.				
OX	NCBI_TaxID-666;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-CLASSICAL INABA 569B;				
RX	MEDLINE-92112300; PubMed-1730472;				
RA	Baudry B., Fasano A., Kelley J., Kaper J.B.;				
RT	"Cloning of a gene (zot) encoding a new toxin produced by Vibrio cholerae.";				
RL	Infect. Immun. 60:428-434(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-KNH002;				
RA	Shin H.J., Park Y.C., Kim Y.C.;				
RT	"Cloning and nucleotide sequence analysis of the virulence gene cassette from Vibrio cholerae KNH002 isolated in Korea.";				
RL	Msalimurhag Holji 35:205-210(1999).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-O139-TOR OGAWA;				
RA	Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.;				
RT	"Cloning and Expression of zot Gene from Vibrio cholerae.";				
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-EL TOR 86015 / SEROTYPE O1;				
RA	Kan B., Liu Y.O., Qi G.M., Gao S.Y.;				
RT	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-EL TOR N16961 / SEROTYPE O1;				
RA	Medline-20406833; PubMed-10952301;				
RT	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,				
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,				
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,				
RA	Emmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,				
RA	McDonald L., Utecherback T., Fleischmann R.D., Nierman W.C., White O.,				
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,				
RA	Fraser C.M.;				
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";				
RL	Nature 406:477-483(2000).				
RN	[6]				
RP	CHARACTERIZATION.				
RX	MEDLINE-91271365; PubMed-2052603;				
RA	Fasano A., Baudry B., Pumpilin D.W., Wasserman S.S., Tall B.D.,				

34	29	72.5	243	1	MOEB_HAEIN	P45211 haemophilus
35	29	72.5	308	1	CPP3_ENTHI	O06964 entamoeba h
36	29	72.5	310	1	ACP2_ENTHI	P36185 entamoeba h
37	29	72.5	315	1	CPPI_ENTHI	O01957 entamoeba h
38	29	72.5	315	1	CPPI_ENTHI	O01957 entamoeba h
39	29	72.5	320	1	VPRT_ASCVP	P36698 apple stem
40	29	72.5	359	1	PEXC_HUMAN	O00623 homo sapien
41	29	72.5	395	1	MAFI_YEAST	P41910 saccharomyc
42	29	72.5	403	1	YCOA_STNP7	P42460 synechococc
43	29	72.5	443	1	AROQ_BACNO	O46550 bacteroides
44	29	72.5	475	1	KPYX_CORGL	O46078 corynebacte
45	29	72.5	482	1	GABD_ECOLI	P25526 escherichia

RA Ketley J.M., Kaper J.B.:  
 RT "Vibrio cholerae produces a second enterotoxin, which affects  
 RL intestinal tight junctions."  
 CC Proc. Natl. Acad. Sci. U.S.A. 88:5242-5246(1991).  
 CC -1- FUNCTION: INCREASES THE PERMEABILITY OF THE SMALL INTESTINE MUCOSA  
 CC BY AFFECTING THE STRUCTURE OF INTERCELLULAR TIGHT JUNCTIONS  
 CC (ZONULA OCCUDENDS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M83563; AAA27582.1; -  
 DR EMBL: AF175708; AAD51358.1; -  
 DR EMBL: AF123049; AAD26854.1; -  
 DR EMBL: AF220606; AAF29547.1; -  
 DR EMBL: AE004224; AAF94615.1; -  
 DR PIR: A43864; A43864.  
 DR TIGR: VC1458; -  
 DR Enterotoxin; Toxin.  
 KM VARIANT 45 45 M -> I (IN STRAIN 569B).  
 FT VARIANT 100 100 V -> A (IN STRAINS 569B AND 86015).  
 FT VARIANT 272 272 V -> A (IN STRAIN 569B).  
 FT VARIANT 281 281 V -> A (IN STRAIN 569B).  
 FT VARIANT 349 349 A -> S (IN STRAIN 86015).  
 FT VARIANT 381 381 K -> R (IN STRAIN 86015).  
 FT CONFLICT 386 399 IKTENDKRGKLSIF -> VKREKESIIKSF (IN REF.  
 FT 4).  
 SQ SEQUENCE 399 AA; 44903 MW; 3C7424B758176774 CRC64;  
 OY 1 GRLVODG 8  
 Db 291 GRLVODG 298  
 Query Match 87.5%; Score 35; DB 1; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 2.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRLVODG 8  
 Db 291 GRLVODG 298  
 RESULT 2  
 Y4GM\_RHISN STANDARD; PRT; 586 AA.  
 ID Y4GM\_RHISN STANDARD; PRT; 586 AA.  
 AC P55469;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROBABLE ABC TRANSPORTER ATP-BINDING PROTEIN Y4GM.  
 GN Y4GM.  
 OS Rhizobium sp. (strain NGR234).  
 OG Plasmid sym pNGR234a.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Frelberg C.A., Frelay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes."  
 RL Nature 387:394-401(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 CC (ABC TRANSPORTERS). MSBA SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000075; AAB91687.1; -  
 DR InterPro: IPR001140; -  
 DR InterPro: IPR001617; -  
 DR Pfam: PF00664; ABC\_membrane; 1.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KM Hypothetical protein; ATP-binding; Transport; Transmembrane; Plasmid.  
 FT TRANSMEM 30 50 POTENTIAL.  
 FT TRANSMEM 66 86 POTENTIAL.  
 FT TRANSMEM 152 172 POTENTIAL.  
 FT TRANSMEM 256 276 POTENTIAL.  
 FT NP\_BIND 379 386 ATP (POTENTIAL).  
 SQ SEQUENCE 586 AA; 64262 MW; 1095DEB82620637 CRC64;  
 OY 1 GRLVODG 8  
 Db 565 GRLVODG 572  
 Query Match 85.0%; Score 34; DB 1; Length 586;  
 Best Local Similarity 87.5%; Pred. No. 6.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRLVODG 8  
 Db 565 GRLVODG 572  
 RESULT 3  
 RPOD\_SYNY3 STANDARD; PRT; 1317 AA.  
 ID RPOD\_SYNY3 STANDARD; PRT; 1317 AA.  
 AC P73334;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE DELTA CHAIN (EC 2.7.7.6).  
 GN RPOC2 OR SLI1789.  
 OS Synecocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
 RA Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions."  
 RL DNA Res. 3:109-136(1996).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: IN CYANOBACTERIA THE RNA POLYMERASE IS COMPOSED OF FOUR  
 CC SUBUNITS: ALPHA, BETA, GAMMA, AND DELTA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D90905; BAA17364.1; -  
 DR InterPro: IPR000722; -  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription.  
 SQ SEQUENCE 1317 AA; 144776 MW; 27B6970469E7E551 CRC64;



```

CC MICROTUBULE ARRAYS IN BOTH MITOSIS AND MEIOSIS.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN KIDNEY, THYMUS AND
CC LIVER. LOW LEVELS IN BRAIN, MUSCLE, LUNG AND HEART.
CC -1- DOMAIN: COMPOSED OF A COILED-COIL CENTRAL REGION FLANKED BY NON-
CC HELICAL N- AND C-TERMINALS.
CC -1- SIMILARITY: STRONG, TO HUMAN KENDRIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05823; AAA17886.1; -.
DR MGD: MGI:102722; Pcnt.
KM Coiled coil; Microtubules.
FT DOMAIN 110 1600 COILED COIL (POTENTIAL).
FT DOMAIN 109 112 POLY-PRO.
FT DOMAIN 383 387 POLY-GIN.
SQ SEQUENCE 1920 AA; 218337 MW; CF1D0ADEC5B73309 CRC64;

Query Match      80.0%; Score 32; DB 1; Length 1920;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8
Db 818 GRLVDEG 825

RESULT 7
DNT1 ALLMI STANDARD; PRT; 101 AA.
AC G9PUD0;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR 1 (FRAGMENT).
GN DMR1.
OS Alligator mississippiensis (American alligator).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Archosauria; Crocodylidae; Alligatorinae; Alligator.
OX NCBI_TaxID=8496;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Embryonic testis;
RX MEDLINE=20067953; PubMed=10604464;
RA Smith C.A., McClive P.J., Western P.S., Reed K.J., Sinclair A.H.;
RT "Conservation of a sex-determining gene.";
RL Nature 402:601-602(1999).
CC -1- FUNCTION: MAY BE REQUIRED FOR TESTIS DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING UROGENITAL SYSTEM
CC (DEVELOPMENTAL STAGES 20 TO 23) AND LATER IN THE GONADS (STAGES 24
CC TO 27) AT BOTH MALE- AND FEMALE-PRODUCING TEMPERATURES, BUT
CC UPREGULATED AT THE MALE-DETERMINING TEMPERATURE. MAY BE HIGHLY
CC EXPRESSED IN TESTICULAR DIFFERENTIATION AND WEAKLY EXPRESSED IN
CC OVARIAN DIFFERENTIATION.
CC -1- SIMILARITY: CONTAINS 1 DM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF192560; AAP03893.1; -.
DR Sexual differentiation; Developmental protein; DNA-binding; Zinc;

```

```

KM Metal-binding; Nuclear protein.
FT NON_TER 1 1 DM.
FT DNA_BIND <1 13
FT NON_TER 101 101
SQ SEQUENCE 101 AA; 10786 MW; 856F8999F0A0980A CRC64;

Query Match      77.5%; Score 31; DB 1; Length 101;
Best Local Similarity 71.4%; Pred. No. 4.1;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLVOD 7
Db 77 GRLVID 83

RESULT 8
PAAY_ECOLI
ID PAAY_ECOLI STANDARD; PRT; 196 AA.
AC P77181; O53020;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PHENYLACETIC ACID DEGRADATION PROTEIN PAAY.
GN PAAY.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubraman S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-W / ATCC 11105;
RA Diaz E.;
RT Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
CC -1- PATHWAY: INVOLVED IN PHENYLACETIC ACID DEGRADATION.
CC -1- SIMILARITY: BELONGS TO THE CYSR/LACA/LPX/NOBL FAMILY OF
CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000237; AAC74482.1; -.
DR EMBL: D90778; BAA15008.1; -.
DR EMBL: X97452; CAA6102.1; -.

```

DR Ecogene: EG13747; paay.  
 DR InterPro: IPR001451; -.  
 DR Pfam: PF00132; hexapep.1.  
 DR PROSITE: PS00101; HEXAPEP\_TRANSFERASES; FALSE\_NEG.  
 KW TRANSFERASE; Repeat.  
 FT VARIANT 75 75 G -> E (IN STRAIN W).  
 FT VARIANT 179 179 I -> V (IN STRAIN W).  
 FT VARIANT 182 182 G -> N (IN STRAIN W).  
 SQ SEQUENCE 196 AA: 21324 MW: FA354F5AA0910DB CRC64:

Query Match 77.5%; Score 31; DB 1; Length 196;  
 Best Local Similarity 62.5%; Pred. No. 8.8;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GRVLQDG 8  
 ||:|:|  
 Db 48 GRIVKDG 55

RESULT 9  
 ID DMT1\_TRASC STANDARD; PRT; 205 AA.  
 AC P57690;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DOUBLESSEX AND MAB-3-RELATED TRANSCRIPTION FACTOR 1 (FRAGMENT).  
 GN DMT1.  
 OS Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Testudines; Cryptodira; Testudinoidae; Emydidae; Trachemys.  
 OX NCBI\_TaxID=34903;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic gonad;  
 RX MEDLINE=20190065; PubMed=10705377;  
 RA Kettlewell J.R., Raymond C.S., Zarkower D.;  
 RT "Temperature-dependent expression of turtle Dmt1 prior to sexual  
 differentiation.";  
 RL Genesis 26:174-178(2000).  
 CC -1- FUNCTION: MAY BE REQUIRED FOR TESTIS DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- DEVELOPMENTAL STAGE: AT STAGES 15, 17 AND 20, EXPRESSION IS HIGHER  
 CC IN THE GENITAL RIDGE/MESONEPHROS COMPLEXES OF EMBRYOS INCUBATED AT  
 CC THE MALE-PROMOTING TEMPERATURE. AT FEMALE PROMOTING TEMPERATURE,  
 CC WEAKLY EXPRESSED IN THE GENITAL RIDGE.  
 CC -1- SIMILARITY: CONTAINS 1 DM DOMAIN.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----

DR EMBL: AF201387; AAF77096.1; -  
 KW Sexual differentiation; Developmental protein; DNA-binding; zinc;  
 KW Metal-binding; Nuclear protein.  
 FT NON\_TER 1 1  
 FT DNA\_BIND <1 29 DM.  
 FT DOMAIN 149 154 POLY-SER.  
 FT NON\_TER 205 205  
 SQ SEQUENCE 205 AA: 22807 MW: 4B3BD9A697AC7B4 CRC64:

Query Match 77.5%; Score 31; DB 1; Length 205;  
 Best Local Similarity 71.4%; Pred. No. 9.2;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GRVLQDG 7  
 ||:|:|

Db 93 GRMLIOD 99

RESULT 10  
 ID DMT1\_CHICK STANDARD; PRT; 311 AA.  
 AC O9PT07; O9PTB5.  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DOUBLESSEX AND MAB-3-RELATED TRANSCRIPTION FACTOR 1 (FRAGMENT).  
 GN DMT1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99178258; PubMed=10080173;  
 RA Nanda I., Shan Z., Scharf M., Burt D.W., Koehler M., Nothwang H.-G.,  
 RA Gruetznher F., Paton I.R., Windsor D., Dunn I., Engel W., Staeheli P.,  
 RA Mizuno S., Haefl T., Schmid M.;  
 RT "300 million years of conserved synteny between chicken 2 and human  
 RT chromosome 9.";  
 RL Nat. Genet. 21:258-259(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=20014847; PubMed=10545231;  
 RA Raymond C.S., Kettlewell J.R., Hirsch B., Bardwell V.J., Zarkower D.;  
 RT "Expression of Dmt1 in the genital ridge of mouse and chicken embryos  
 RT suggests a role in vertebrate sexual development.";  
 RL Dev. Biol. 215:208-220(1999).  
 CC -1- FUNCTION: MAY BE REQUIRED FOR TESTIS DEVELOPMENT. TWO COPIES OF  
 CC DMT1 ARE PROBABLY REQUIRED FOR TESTIS FORMATION, WHEREAS A SINGLE  
 CC COPY ALONG WITH THE W CHROMOSOME LEADS TO FEMALE SEXUAL  
 CC DIFFERENTIATION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT STAGE 19 BEFORE SEXUAL  
 CC DIFFERENTIATION, AT STAGE 25 WHEN THE GENITAL RIDGES ARE FORMING  
 CC AND ALSO AT STAGE 31 WHEN SEXUAL DIFFERENTIATION IS BEGINNING.  
 CC ALSO EXPRESSED FROM STAGE 25 IN THE WOLFPIAN DUCTS. MORE ABUNDANT  
 CC IN Z2 (MALE) THAN ZW (FEMALE) GONADS.  
 CC -1- SIMILARITY: CONTAINS 1 DM DOMAIN.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----

DR EMBL: AF211349; AAF19034.1; -  
 DR EMBL: AF123456; AAF19666.1; -  
 DR InterPro: IPR001273; -  
 DR Pfam: PF00751; DM-domain.1;  
 KW Sexual differentiation; Developmental protein; DNA-binding; zinc;  
 KW Metal-binding; Nuclear protein.  
 FT NON\_TER 1 1  
 FT DNA\_BIND 9 55 DM.  
 FT DOMAIN 87 92 POLY-SER.  
 FT DOMAIN 109 115 POLY-ALA.  
 FT DOMAIN 177 180 POLY-SER.  
 FT DOMAIN 258 286 PRO/SER-RICH.  
 FT CONFLICT 2 2  
 FT CONFLICT 54 54 A -> G (IN REF. 2).  
 FT CONFLICT 54 54 V -> A (IN REF. 2).  
 SQ SEQUENCE 311 AA: 33730 MW: 326FEFF9335E2631 CRC64:

Query Match 77.5%; Score 31; DB 1; Length 311;  
 Best Local Similarity 71.4%; Pred. No. 15;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVOD 7  
11:111  
Db 119 GRMLIOD 125

RESULT 11  
IADA\_ECOLI  
ID IADA\_ECOLI STANDARD; PRT; 390 AA.  
AC P9377;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ISOASPARTYL DIPEPTIDASE (EC 3.4.19.-).  
GN IADA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.; PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RC STRAIN-K12 / W3110;  
RX MEDLINE=95181377; PubMed=7876157;  
RA Gary J.D., Clarke S.;  
RT "Purification and characterization of an isoaspartyl dipeptidase from  
RT Escherichia coli.";  
RT J. Biol. Chem. 270:4076-4087(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=9534362; PubMed=7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RA Baitner F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes.";  
RL Nucleic Acids Res. 23:2105-2119(1995).  
CC -1- FUNCTION: CATALYZES THE HYDROLYTIC CLEAVAGE OF A SUBSET OF L-  
CC ISOASPARTYL (L-BETA-ASPARTYL) DIPEPTIDES. USED TO DEGRADE PROTEINS  
CC DAMAGED BY L-ISOASPARTYL RESIDUES FORMATION.  
CC -1- SUBCELLULAR LOCATION: CYTOSOL; CYTOSOL.  
CC -1- SIMILARITY: SOME, IN THE N-TERMINAL, TO DIHYDROLYASE.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M38.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M38.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U15029; AAC43299.1; -.  
DR EMBL: U14003; AAA97224.1; -.  
DR EMBL: AE000503; AAC77284.1; -.  
DR MEROPS: M38.001; -.  
DR EcGene: EG12567; iada.  
KW Hydrolyase; Metalloprotease.  
SQ SEQUENCE 390 AA; 41084 MW; 9CECC838381545B5 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 390;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVODG 8  
11:11111  
Db 372 GKILVODG 379

RESULT 12  
YF03\_METTH  
ID YF03\_METTH STANDARD; PRT; 427 AA.

AC 027549;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL PROTEIN MTH1505.  
GN MTH1505.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DELTA H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7153(1997).  
CC -1- SIMILARITY: BELONGS TO THE ATZ/TRZ FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AE000910; AAB85980.1; -.  
DR InterPro: IPR002604; -.  
DR Pfam: PF01685; ATZ-TRZ; 1.  
KW Hypothetical protein; Hydrolase.  
SQ SEQUENCE 427 AA; 46441 MW; 439527DD142182CD CRC64;

Query Match 77.5%; Score 31; DB 1; Length 427;  
Best Local Similarity 62.5%; Pred. No. 21;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVODG 8  
11:11111  
Db 397 GRLLVODG 404

RESULT 13  
WCAM\_SALTY  
ID WCAM\_SALTY STANDARD; PRT; 467 AA.  
AC P26389;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE COLANIC ACID BIOSYNTHESIS PROTEIN WCAM.  
GN WCAM.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LT2;  
RX MEDLINE=91260454; PubMed=1710759;  
RA Jiang X.-M., Neal B., Santiago F., Lee S.-J., Romana L.K., Reeves P.R.;  
RT "Structure and sequence of the rfb (O antigen) gene cluster of  
RT Salmonella serovar typhimurium (strain LT2).";  
RL Mol. Microbiol. 5:693-713(1991).  
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE  
CC COLANIC ACID.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X56793; CAA40113.1; -  
 CC PIR: S15297; S15297.  
 CC StryGene: SG10448; wcam.  
 CC LipoPolysaccharide biosynthesis.  
 CC SEQUENCE 467 AA; 50958 MW; 9DCCFD551218E6E8 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 467;  
 Best Local Similarity 62.5%; Pred. No. 24;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GRLVVDG 8  
 Db 97 GRLVVDG 104

RESULT 14  
 MENG\_ECOLI STANDARD; PRT; 161 AA.  
 AC P32165;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE S-ADENOSYLMETHYLTRANSFERASE  
 DE (EC 2.1.1.-).  
 GN MENG.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA Hudspeth M.E.S., Suvarna K., Meganathan R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=93347969; PubMed=8346018;  
 RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the  
 RT region from 87.2 to 89.2 minutes."  
 RL Nucleic Acids Res. 21:3391-3398(1993).  
 CC -1- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK).  
 CC -1- PATHWAY: LAST STEP IN MENAQUINONE BIOSYNTHESIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U56082; AAB01208.1; -  
 CC EMBL: L19201; AAB03061.1; -  
 CC EMBL: AE000467; AAC76911.1; -  
 CC PIR: S40872; S40872.  
 CC Ecogene: EG11879; meng.  
 CC Menadiquinone biosynthesis; Transferrase; Methyltransferase.  
 CC SEQUENCE 161 AA; 17360 MW; B30371B838DE21F8 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 161;

Best Local Similarity 75.0%; Pred. No. 12;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRLVVDG 8  
 Db 58 GRLVVDG 65

RESULT 15  
 YCBO\_ECOLI STANDARD; PRT; 182 AA.  
 AC P75855;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL FIMBRIAL-LIKE PROTEIN IN PEPN-PYRD INTERGENIC REGION  
 DE PRECURSOR.  
 GN YCBO.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map."  
 RL DNA Res. 3:137-155(1996).  
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000196; AAC74024.1; -  
 CC EMBL: D90732; BAA35693.1; -  
 CC Ecogene: EG13709; ycbo.  
 CC InterPro: IPR000259; -  
 CC Pfam: PF00419; Fimbril; 1.  
 CC Hypothetical protein: Fimbril; Signal.  
 FT SIGNAL  
 FT CHAIN 1 24 POTENTIAL.  
 FT CHAIN 25 182 HYPOTHETICAL FIMBRIAL-LIKE PROTEIN YCBO.  
 CC SEQUENCE 182 AA; 19025 MW; A984F5CB2B6E86BC CRC64;

Query Match 75.0%; Score 30; DB 1; Length 182;  
 Best Local Similarity 62.5%; Pred. No. 14;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GRLVVDG 8  
 Db 143 GRLVVDG 150

Wed Jun 13 15:01:22 2001

pct-us01-05825a-8.rsp

Page 8

Search completed: June 13, 2001, 14:21:43  
Job time: 803 sec

---





DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
 DE ZONULAR OCCLUDENS TOXIN (ZONA OCCLUDENS TOXIN).  
 GN ZOT OR VC1458.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 ON NCBI\_TaxID=666;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KNH002;  
 RA Shin H.J., Park Y.C., Kim Y.C.;  
 RT "Cloning and nucleotide sequence analysis of the virulence gene  
 RT cassette from Vibrio cholerae KNH002 isolated in Korea."  
 RL Misaiimuhag Holji 35:205-210(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0139-TOR OGAWA;  
 RA Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.;  
 RT "Cloning and Expression of zot Gene from Vibrio cholerae."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA McAnulla M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,  
 RA Emondalad L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae."  
 RL Nature 406:477-483(2000).  
 DR EMBL: AF175708; AAB51358.1; -  
 DR EMBL: AF123049; AAD26854.1; -  
 DR EMBL: AE004224; AAF94615.1; -  
 DR TIGR: VC1458; -  
 SQ SEQUENCE 399 AA; 44903 MW; 3C7424B758176774 CRC64;

Query Match 87.5%; Score 35; DB 2; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 Db 291 GRLVODG 298

RESULT 3  
 O9L706 PRELIMINARY; PRT; 399 AA.  
 AC O9L706.  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
 DE ZOT.  
 GN ZOT.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 ON NCBI\_TaxID=666;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-86015;  
 RA Kan B., Liu Y.Q., Ol G.M., Gao S.Y.;  
 RT "Vibrio cholerae nct-cxphr whole genome, include rstr(Rstr),  
 RT rstr(Rstr), rstr(Rstr), cep(Cep), orfU(OrfU), ace(Ace) and zot(Zot)  
 RT genes."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF220606; AAF29547.1; -  
 GN F21M12.28.  
 SQ SEQUENCE 399 AA; 44990 MW; CF6A3DBC9E23EE1 CRC64;

Query Match 87.5%; Score 35; DB 2; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 Db 291 GRLVODG 298

RESULT 4  
 P91352 PRELIMINARY; PRT; 687 AA.  
 AC P91352;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, last annotation update)  
 DE CODED FOR BY C. ELEGANS CDNA CM9G4.  
 GN F56A3.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 ON NCBI\_TaxID=6239;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsy T., Cooper J., Coulson A.,  
 RA Craxton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Geisel C., Bradshaw H., Wamsley P.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U80446; AAB37805.1; -  
 SQ SEQUENCE 687 AA; 78247 MW; C4FP64BA73DD92BE CRC64;

Query Match 87.5%; Score 35; DB 5; Length 687;  
 Best Local Similarity 87.5%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
 Db 587 GRLVODG 594

RESULT 5  
 O04511 PRELIMINARY; PRT; 447 AA.  
 AC O04511;  
 DT 01-JUL-1997 (Tremblrel. 04, Created)  
 DT 01-JUL-1997 (Tremblrel. 04, last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, last annotation update)  
 DE SEQUENCE OF BAC F21M12 FROM ARABIDOPSIS THALIANA CHROMOSOME 1,  
 DE COMPLETE SEQUENCE.  
 GN F21M12.28.  
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:  
 OC Magnoliophyta: eudicotyledons, core eudicots: Rosidae: eurosids II;  
 OC Brassicales: Brassicaceae: Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Vysotskaya V.S., Osborne B.I., Toriumi M., Yu G., Oji O., Shen Y.K.,  
 RA Araujo R., Au M., Buehler E., Conway A.B., Conway A.R., Dewar K.,  
 RA Feng J., Kim C., Kurtz D., Li Y., Shinn P., Sun H., Davis R.W.,  
 RA Ecker J.R., Federspiel N.A., Theologis A.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC000132; AAB60742.1; -;  
 DR MENDEL: 17346; Arath:2796:17346.  
 SQ SEQUENCE 447 AA; 51770 MW; 010BF7D3342331F CRC64;

Query Match 85.0%; Score 34; DB 10; Length 447;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOD 7  
 Db 309 GRLLVOD 315

RESULT 6  
 ID P96438 PRELIMINARY; PRT; 473 AA.  
 AC P96438;  
 DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE EXP2.  
 OS Sinorhizobium meliloti.  
 OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;  
 OC Rhizobiaceae: Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RM2011;  
 RA MEDLINE-97175570; PubMed-9023225;  
 RA Becker A., Rueberg H., Walker G.C., Puhler A.,  
 RA Ivashina T., Cheng H., Kuester H., Roxlau A.A., Keller M.,  
 RA "The 32-kilobase exp gene cluster of Rhizobium meliloti directing the  
 RT biosynthesis of galactoglucan: genetic organization and properties of  
 RT the encoded gene products."  
 RL J. Bacteriol. 179:1375-1384(1997).  
 DR EMBL: Z79692; CAB01939.1; -;  
 DR INTERPRO: IPR002215; -;  
 DR PFAM: PF00529; HLYD; 1;  
 SQ SEQUENCE 473 AA; 51479 MW; C0C3B168D9C23D85 CRC64;

Query Match 85.0%; Score 34; DB 2; Length 473;  
 Best Local Similarity 75.0%; Pred. No. 56;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOD 8  
 Db 107 GRLLVOD 114

RESULT 7  
 ID O9K8J8 PRELIMINARY; PRT; 382 AA.  
 AC O9K8J8;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE BH3008 PROTEIN.  
 GN BH3008.  
 OS Bacillus halodurans.

OC Bacteria: Firmicutes: Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=8665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C-125 / JCM 9153;  
 RA Takami H., Nakasone K., Takaki Y.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF001517; BAB06727.1; -;  
 SQ SEQUENCE 382 AA; 41365 MW; B6B54AF6BC163565 CRC64;

Query Match 82.5%; Score 33; DB 2; Length 382;  
 Best Local Similarity 75.0%; Pred. No. 71;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOD 8  
 Db 20 GRLLVOD 27

RESULT 8  
 ID O29120 PRELIMINARY; PRT; 482 AA.  
 AC O29120;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE 4-HYDROXYBUTYRATE COA TRANSFERASE (CAR2-1).  
 GN AF1145.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota: Archaeoglobales: Archaeoglobaceae;  
 OC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE-98049343; PubMed-9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kervilave A.R., Graham D.E., Kyriades N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Ketch C.I., McNeil L.K., Badger J.H., Glodex A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Arlatch P., Kaine B.P., Sykes S.M.,  
 RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus."  
 RL Nature 390:364-370(1997).  
 DR EMBL: AE001025; AAB90101.1; -;  
 DR TIGR: AF1145; -;  
 DR PRODOM: PD005956; -; 1.  
 KW Hypothetical protein; Transferase.  
 SQ SEQUENCE 482 AA; 54018 MW; 895ADBD410E1CAF8 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 482;  
 Best Local Similarity 75.0%; Pred. No. 92;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOD 8  
 Db 225 GRLLVOD 232

RESULT 9  
 ID O55897 PRELIMINARY; PRT; 495 AA.  
 AC O55897;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE HYDROTHERMAL 53.0 KDA PROTEIN.  
 GN SLR0121.  
 OS *Synechocystis* sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Maraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL: D64005; BAA10659.1; -  
 DR INTERPRO: IPR001466; -  
 DR PFM: PF00144; beta-lactamase; 1.  
 DR Hypothetical protein.  
 KW SEQUENCE 495 AA; 52995 MW; 732AE86B920978B5 CRC64;  
 SQ  
 Query Match 82.5%; Score 33; DB 2; Length 495;  
 Best Local Similarity 85.7%; Pred. No. 95;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GRLVOD 7  
 Db 451 GRLVOD 457  
 RESULT 10  
 ID 09VBN8 PRELIMINARY; PRT: 111 AA.  
 AC 09VBN8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE CG5039 PROTEIN.  
 GN CG5039.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta.  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; *Drosophilidae*; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ideyama C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclebo J.M.,  
 RA Palazzolo R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,  
 RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003753; AAF56492.1; -  
 DR FLYBASE: FBgn0039356; CG5039.  
 DR SEQUENCE 111 AA; 12260 MW; 6427966E3D467045 CRC64;  
 SQ  
 Query Match 80.0%; Score 32; DB 5; Length 111;  
 Best Local Similarity 75.0%; Pred. No. 30;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRLVODG 8  
 Db 97 GRLVODG 104  
 RESULT 11  
 ID 09UOYO PRELIMINARY; PRT: 365 AA.  
 AC 09UOYO;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE O-METHYLTRANSFERASE.  
 GN MT-1.  
 OS *Aspergillus parasiticus*.  
 OC Eukaryota; Fungi; Ascomycota; Eurotiiales; *Trichocomaceae*;  
 OC anamorphic *Trichocomaceae*; *Aspergillus*.  
 OX NCBI\_TaxID=5067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NIH-26;  
 RA Motomura M., Chihaya N., Shinozawa T., Hamasaki T., Yabe K.;  
 RT "Cloning and characterization of the O-methyltransferase I gene (*omtA*)  
 RT from *Aspergillus parasiticus* associated with the conversions of  
 RT demethylsterigmatocystin to sterigmatocystin and  
 RT dihydrodemethylsterigmatocystin to dihydrosterigmatocystin in  
 RT aflatoxin biosynthesis.";  
 RL Appl. Environ. Microbiol. 65:4987-4994(1999).  
 DR EMBL: AB022906; BAA6104.1; -  
 DR EMBL: AB022905; BAA6103.1; -  
 DR INTERPRO: IPR000183; -  
 DR INTERPRO: IPR001077; -  
 DR INTERPRO: IPR001601; -  
 DR PFM: PF00891; Methyltransf\_2; 1.  
 DR PROSITE: PS00879; ODR DC 2.2; UNKNOWN\_1.  
 DR Transferase; Methyltransferase.

FT INIT\_MET 0  
SQ SEQUENCE 385 AA; 43023 MW; C9866E7A96272005 CRC64;

Query Match  
Best Local Similarity 80.0%; Score 32; DB 3; Length 385;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOD 7  
DB 246 GRLLVOD 252

RESULT 12

ID O9UV62 PRELIMINARY; PRT: 386 AA.  
AC O9UV62;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE O-METHYLTRANSFERASE B.  
GN OMTB.  
OS Aspergillus parasiticus.  
OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;  
OC anamorphic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5067;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SU-1;  
RA Yu J., Molosnuk C.P., Bhattacharjee D., Cleveland T.E.;  
RT "Characterization of avfa and omtb genes involved in aflatoxin biosynthesis";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154050; AAF25603.1; -  
DR INTERPRO: IPR001073; -  
DR INTERPRO: IPR001077; -  
DR INTERPRO: IPR001601; -  
DR PFAM: PF00891; Methyltransferase\_2; 1.  
DR PROSITE: PS00879; ODR\_DC\_2; UNKNOWN\_1.  
KW Transferase; Methyltransferase.  
SQ SEQUENCE 386 AA; 43142 MW; 13D6F1ED63CF663D CRC64;

Query Match  
Best Local Similarity 80.0%; Score 32; DB 3; Length 386;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOD 7  
DB 247 GRLLVOD 253

RESULT 13

ID O9P900 PRELIMINARY; PRT: 386 AA.  
AC O9P900;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE O-METHYLTRANSFERASE B.  
GN OMTB.  
OS Aspergillus flavus.  
OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;  
OC anamorphic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5059;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CRA01-2B;  
RA Yu J., Molosnuk C.P., Bhattacharjee D., Cleveland T.E.;  
RT "Cloning and characterization of avfa and omtb genes involved in aflatoxin biosynthesis in three Aspergillus species";  
RL Gene 248:157-167(2000).  
DR EMBL; AF159789; AAF26223.1; -

KW Transferase; Methyltransferase.  
SQ SEQUENCE 386 AA; 43160 MW; F199017CC28DE9C1 CRC64;

Query Match  
Best Local Similarity 80.0%; Score 32; DB 3; Length 386;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLLVOD 7  
DB 247 GRLLVOD 253

RESULT 14

ID O95072 PRELIMINARY; PRT: 547 AA.  
AC O95072;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE HHR21SPB (FRAGMENT).  
GN HHR21SPB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA McKay M.J., van der Spek P., Kanaar R., Bootsma D., Hoeijmakers J.H.;  
RT "HHR21SPB, a second human gene homologous to the rad21 subtelomeric repeats of the pombe DNA double strand break repair gene";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF006264; AAD01193.1; -  
FT NON\_TER 547  
SQ SEQUENCE 547 AA; 62613 MW; 544AA31FD038D7B CRC64;

Query Match  
Best Local Similarity 80.0%; Score 32; DB 4; Length 547;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVOD 8  
DB 536 GRLLVOD 543

RESULT 15

ID O9NV09 PRELIMINARY; PRT: 547 AA.  
AC O9NV09;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE CDNA FLJ10573 FIS, CLONE NT2RP2003177.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakematsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.;  
RT "NED0 human cDNA sequencing project";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK001435; BAA91690.1; -  
SQ SEQUENCE 547 AA; 62585 MW; 05A1CFC874F6C3BB CRC64;

Query Match  
Best Local Similarity 80.0%; Score 32; DB 4; Length 547;

Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLLYODG 8

Db 536 GRLLIOPG 543

Search completed: June 13, 2001, 14:20:28  
Job time: 728 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:33 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-8

Perfect score: 40

Sequence: 1 GRLLVODG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents: AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCYUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	87.5	400	2	US-08-624-601-8
2	32	80.0	2391	2	US-08-446-855A-2
3	32	80.0	2391	4	US-09-150-741-2
4	30	75.0	117	1	US-08-499-215-3
5	30	75.0	312	4	US-08-460-900C-14
6	30	75.0	313	2	US-08-356-060A-14
7	30	75.0	456	2	US-08-709-979A-1
8	30	75.0	456	4	US-08-709-974A-11
9	29	72.5	27	2	US-08-709-874A-1
10	29	72.5	272	2	US-08-709-874A-2
11	29	72.5	299	2	US-08-709-874A-19
12	29	72.5	375	1	US-07-817-917-1
13	29	72.5	425	3	US-08-946-329A-19
14	29	72.5	425	4	US-08-567-357A-19
15	29	72.5	425	5	US-08-567-357A-19
16	29	72.5	425	5	PCT-US95-15463-19
17	29	72.5	428	4	US-09-329-350-33
18	29	72.5	443	4	US-08-476-008-69
19	29	72.5	443	1	US-08-306-063-69
20	29	72.5	443	1	US-08-833-485-69
21	29	72.5	596	2	US-08-392-806A-2
22	29	72.5	606	2	US-08-392-806A-6
23	29	72.5	659	2	US-08-709-874A-20
24	29	72.5	2154	2	US-08-841-349-4
25	28	70.0	118	2	US-08-652-816A-12
26	28	70.0	128	2	US-08-318-837-51
27	28	70.0	247	3	US-08-619-812-2

28	28	70.0	311	2	US-08-318-837-7	Sequence 7, Appl
29	28	70.0	341	1	US-08-190-802A-45	Sequence 45, Appl
30	28	70.0	409	3	US-08-807-342B-8	Sequence 8, Appl
31	28	70.0	498	3	US-08-889-841-2	Sequence 2, Appl
32	28	70.0	532	4	US-09-008-271A-12	Sequence 12, Appl
33	28	70.0	600	2	US-08-392-806A-4	Sequence 4, Appl
34	28	70.0	2860	2	US-08-826-267-2	Sequence 2, Appl
35	27	67.5	79	4	US-09-025-151-7	Sequence 7, Appl
36	27	67.5	310	2	US-08-484-956-88	Sequence 88, Appl
37	27	67.5	310	2	US-08-757-653-88	Sequence 88, Appl
38	27	67.5	311	1	US-08-872-784-1	Sequence 1, Appl
39	27	67.5	311	2	US-09-100-851-1	Sequence 1, Appl
40	27	67.5	311	4	US-09-265-294-1	Sequence 1, Appl
41	27	67.5	315	2	US-08-484-956-91	Sequence 91, Appl
42	27	67.5	315	2	US-08-757-653-91	Sequence 91, Appl
43	27	67.5	320	2	US-08-757-653-163	Sequence 163, App
44	27	67.5	320	2	US-08-823-516-61	Sequence 61, Appl
45	27	67.5	320	3	US-08-759-038-102	Sequence 102, App

#### ALIGNMENTS

RESULT 1  
US-08-624-601-8  
; Sequence 8, Application US/08624601  
; Patent No. 5882653  
; GENERAL INFORMATION:  
; APPLICANT: Kaper Dr., James B.  
; APPLICANT: Levine Dr., Myron M.  
; TITLE OF INVENTION: Vibrio cholerae O1 (CVD11) and non-O1  
; TITLE OF INVENTION: CVD112 and CVD112RM, serogroup vaccine strains, methods  
; TITLE OF INVENTION: of making same and products thereof  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Spencer & Frank  
; STREET: 1100 New York Ave. N.W. Suite 300 East  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/624,601  
; FILING DATE: 08-APR-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Schellier Dr., John W.  
; REGISTRATION NUMBER: 26,031  
; REFERENCE/DOCKET NUMBER: BANC20019P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)414-4000  
; TELEFAX: (202)414-4040  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 400 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Vibrio cholerae  
; STRAIN: El Tor 7946  
; IMMEDIATE SOURCE:  
; CLONE: zot  
US-08-624-601-8

Query Match 87.5%; Score 35; DB 2; Length 400;  
Best Local Similarity 87.5%; Pred. No. 7.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
|||:|  
Db 291 GRLVODG 298

RESULT 2  
US-08-446-855A-2  
; Sequence 2, Application US/08446855A  
; Patent No. 5849573  
; GENERAL INFORMATION:  
; APPLICANT: Stewart, Thomas S  
; APPLICANT: Flores, Maria V  
; APPLICANT: O'Sullivan, William J  
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderhye PC  
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,855A  
; FILING DATE: 06-Jul-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mitchard, Leonard C  
; REGISTRATION NUMBER: 29,009  
; REFERENCE/DOCKET NUMBER: 47-80  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2391 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-446-855A-2

Query Match 80.0%; Score 32; DB 2; Length 2391;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
|||:|  
Db 29 GRLLEDG 36

RESULT 3  
US-09-150-741-2  
; Sequence 2, Application US/09150741  
; Patent No. 6183996  
; GENERAL INFORMATION:  
; APPLICANT: Stewart et al.  
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
; Patent No. 6183996  
; TITLE OF INVENTION: Synthetase II  
; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/09/150,741  
; EARLIER FILING DATE: 1998-09-10  
; EARLIER APPLICATION NUMBER: PL6380  
; EARLIER FILING DATE: 1992-12-16  
; EARLIER APPLICATION NUMBER: AU93/00617  
; EARLIER FILING DATE: 1993-12-02  
; EARLIER APPLICATION NUMBER: 08/446,855  
; EARLIER FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2391  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: protein  
US-09-150-741-2

Query Match 80.0%; Score 32; DB 4; Length 2391;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
|||:|  
Db 29 GRLLEDG 36

RESULT 4  
US-08-499-215-3  
; Sequence 3, Application US/08499215  
; Patent No. 5612204  
; GENERAL INFORMATION:  
; APPLICANT: Saeki, Hisashi  
; APPLICANT: Miura, Akira  
; TITLE OF INVENTION: BIOLOGICAL DEGRADATIVE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W.  
; STATE: Washington D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; SOFTWARE: 6 WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,215  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP Hei-6-179689  
; FILING DATE: 08-JUL-1994  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-499-215-3

Query Match 75.0%; Score 30; DB 1; Length 117;  
Best Local Similarity 62.5%; Pred. No. 21;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVODG 8  
|||:|



Db 95 GRIVSDG 102

## RESULT 5

US-08-460-900C-14

Sequence 14, Application US/08460900C

Patent No. 6165747

## GENERAL INFORMATION:

APPLICANT: Ingham, Phillip W.

APPLICANT: McMahon, Andrew P.

APPLICANT: Tablin, Clifford J.

APPLICANT: Bumcrot, David A.

APPLICANT: Marti-Gorosoziza, Elisa

TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing

TITLE OF INVENTION: Proteins and Uses Related Thereto

NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY, HOAG &amp; ELIOT LLP

STREET: One Post Office Square

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,900C

FILING DATE: 5-JUNE-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/435,093

FILING DATE: 4-MAY-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/356,060

FILING DATE: 14-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/176,427

FILING DATE: 30-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.

REGISTRATION NUMBER: 36,709

REFERENCE/DOCKET NUMBER: HMV-006.05

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 832-7000

TELEFAX: (617) 832-7000

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 312 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-460-900C-14

Query Match 75.0%; Score 30; DB 4; Length 312;

Best Local Similarity 62.5%; Pred. No. 61;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVODG 8

Db 293 GRLLBEG 300

## RESULT 6

US-08-356-060A-14

Sequence 14, Application US/08356060A

Patent No. 5844079

## GENERAL INFORMATION:

APPLICANT: Ingham, Phillip W.

APPLICANT: McMahon, Andrew P.

APPLICANT: Tablin, Clifford J.

TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing

TITLE OF INVENTION: Proteins and Uses Related Thereto

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE &amp; COCKFIELD

STREET: 60 State Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII(text)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/356,060A

FILING DATE: 14-DEC-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/176,427

FILING DATE: 30-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.

REGISTRATION NUMBER: 36,709

REFERENCE/DOCKET NUMBER: HMT-006CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 313 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-356-060A-14

Query Match 75.0%; Score 30; DB 2; Length 313;

Best Local Similarity 62.5%; Pred. No. 62;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLVODG 8

Db 293 GRLLBEG 300

## RESULT 7

US-08-709-979A-1

Sequence 1, Application US/08709979A

Patent No. 5912157

GENERAL INFORMATION:

APPLICANT: Claus von der Osten

APPLICANT: Martin Sch lein

TITLE OF INVENTION: No. 5912157el Alkaline Cellulases

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5912157o No. 5912157disk of No. 5912157th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/709,979A

FILING DATE: 09-SEP-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4160, 404-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 456 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-709-974A-1

Query Match 75.0%; Score 30; DB 2; Length 456;  
Best Local Similarity 85.7%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RLIVQDG 8  
|||  
Db 305 RLIVQDG 311

RESULT 8  
US-08-709-974A-1  
Sequence 11, Application US/08709974A  
Patent No. 6117664  
GENERAL INFORMATION:  
APPLICANT: Sch Jain, Martin  
APPLICANT: Rosholt, Peter  
APPLICANT: Nielsen, Jack Bech  
APPLICANT: Hansen, Svend Aage  
APPLICANT: von der Osten, Claus  
TITLE OF INVENTION: No. 6117664el Alkaline Cellulases  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 61176640 No. 6117664disk of No. 6117664th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,974A  
FILING DATE: 09-SEP-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta  
REGISTRATION NUMBER: 35,127Y  
REFERENCE/DOCKET NUMBER: 4160, 414-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 456 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-709-974A-11

Query Match 75.0%; Score 30; DB 4; Length 456;  
Best Local Similarity 85.7%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RLIVQDG 8  
|||  
Db 305 RLIVQDG 311

RESULT 9  
US-08-709-874A-1  
Sequence 1, Application US/08709874A  
Patent No. 5854040  
GENERAL INFORMATION:  
APPLICANT: Ozaki, Akio  
APPLICANT: Mori, Hideo  
APPLICANT: Shibasaki, Takeshi  
APPLICANT: Ando, Katsuhiko  
APPLICANT: Chiba, Shigeru  
TITLE OF INVENTION: Process for Producing  
TRANS-4-HYDROXY-L-PROLINE  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ANTONELLI, TERRY, STOUT AND KRAUS, LLP  
STREET: 1300 NORTH SEVENTEENTH STREET  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22209  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,874A  
FILING DATE: 09-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/301,653  
FILING DATE: 07-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/482,554  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Teity, David T.  
REGISTRATION NUMBER: 20178  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-312-6600  
TELEFAX: 703-312-6666  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Dactylosporangium sp.  
STRAIN: RH1  
US-08-709-874A-1

Query Match 72.5%; Score 29; DB 2; Length 27;  
Best Local Similarity 62.5%; Pred. No. 6.8;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLLVQDG 8  
|||  
Db 14 GYLLIEDG 21

RESULT 10  
US-08-709-874A-2

```
; Sequence 2, Application US/08709874A
; Patent No. 5854040
; GENERAL INFORMATION:
; APPLICANT: Ozaki, Akio
; APPLICANT: Mori, Hideo
; APPLICANT: Shibasaki, Takeshi
; APPLICANT: Ando, Katsuhiko
; APPLICANT: Chiba, Shigeru
; TITLE OF INVENTION: Process for Producing
; TITLE OF INVENTION: Trans-4-Hydroxy-L-Proline
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANTONELLI, TERRY, STOUT AND KRAUS, LLP
; STREET: 1300 NORTH SEVENTEENTH STREET
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22209
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,874A
; FILING DATE: 09-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/301,653
; FILING DATE: 07-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/482,554
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Terry, David T.
; REGISTRATION NUMBER: 20178
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-312-6600
; TELEFAX: 703-312-6666
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Dactylosporangium sp.
; STRAIN: RH1
; US-08-709-874A-2

Query Match          72.5%; Score 29; DB 2; Length 272;
Best Local Similarity 62.5%; Pred. No. 86;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
; TITLE OF INVENTION: Process for Producing
; TITLE OF INVENTION: Trans-4-Hydroxy-L-Proline
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANTONELLI, TERRY, STOUT AND KRAUS, LLP
; STREET: 1300 NORTH SEVENTEENTH STREET
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22209
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,874A
; FILING DATE: 09-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/301,653
; FILING DATE: 07-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/482,554
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Terry, David T.
; REGISTRATION NUMBER: 20178
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-312-6600
; TELEFAX: 703-312-6666
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Dactylosporangium sp.
; STRAIN: RH1
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 35 to 299
; IDENTIFICATION METHOD: by similarity with known sequence
; IDENTIFICATION METHOD: or to an established consensus
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE: pBluescriptIIKS+
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1 to 34
; IDENTIFICATION METHOD: by similarity with known sequence
; IDENTIFICATION METHOD: or to an established consensus
; US-08-709-874A-19

Query Match          72.5%; Score 29; DB 2; Length 299;
Best Local Similarity 62.5%; Pred. No. 95;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

APPLICANT: Masako NISHIZAWA  
APPLICANT: Hideo OKAWA  
APPLICANT: Masaharu KANAOKA  
TITLE OF INVENTION: GENE ENCODING ASYMMETRICALLY ACTIVE  
TITLE OF INVENTION: ESTERASE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 540563th Washington Street  
CITY: Falls Church  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22046-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 Inch, 500 KB  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/817,917  
FILING DATE: 19920109  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 001537/1991  
FILING DATE: 10-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 20-2941P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-241-1300  
TELEFAX: 703-241-2848  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Arthrobacter globiformis  
STRAIN: SC-6-98-28 (Ferm BP-3618)  
US-07-817-917-1

Query Match 72.5% Score 29; DB 1; Length 375;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLYVDG 8  
|||||  
Db 71 LLYVDG 76

RESULT 13  
US-08-946-329A-19  
Sequence 19, Application US/08946329A  
Patent No. 6057091  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.  
APPLICANT: Porter, Jeffrey A.  
TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,329A  
FILING DATE: 07-OCT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/061,323  
FILING DATE: 07-OCT-1996  
APPLICATION NUMBER: 08/729,743  
FILING DATE: 10-JUL-1996  
APPLICATION NUMBER: 08/567,357  
FILING DATE: 04-DEC-1995  
APPLICATION NUMBER: 08/349,498  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/140001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 425 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: protein  
US-08-946-329A-19

Query Match 72.5% Score 29; DB 3; Length 425;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVOD 7  
|||||  
Db 233 GRLVOD 239

RESULT 14  
US-08-567-357A-19  
Sequence 19, Application US/08567357A  
Patent No. 6132728  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.  
APPLICANT: Moon, Randall T.  
APPLICANT: Porter, Jeffrey A.  
TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/567,357A  
FILING DATE: 04-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/349,498  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/080001  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 425 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: protein  
US-08-567-357A-19

Query Match 72.5% Score 29; DB 4; Length 425;  
Best Local Similarity 85.7% Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVOD 7  
|||||  
Db 233 GRLVSD 239

RESULT 15  
PCT-US95-15463-19  
Sequence 19, Application PC/TUS9515463  
GENERAL INFORMATION:  
APPLICANT: The Johns Hopkins University School of Medicine  
TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/15463  
FILING DATE: 01-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/080W01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 425 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: both  
MOLECULE TYPE: protein  
PCT-US95-15463-19

Query Match 72.5% Score 29; DB 5; Length 425;  
Best Local Similarity 85.7% Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVOD 7  
|||||  
Db 233 GRLVSD 239

Search completed: June 13, 2001, 14:16:33  
Job time: 494 sec

*This Page Blank (usp10)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:34 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-9  
Perfect score: 48  
Sequence: 1 GRGCVGPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues  
Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_0401:\*

- 1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 21: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*
- 22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	8	21	V79113
2	41	85.4	8	21	V79125
3	40	83.3	8	21	V79114
4	39	81.2	8	21	V79105
5	38	79.2	8	21	V79109
6	38	79.2	8	21	V79115
7	38	79.2	38	17	R91407
8	38	79.2	38	17	R91405
9	38	79.2	38	17	R91405
10	38	79.2	38	17	R91404
11	38	79.2	38	17	R91403

12	38	79.2	42	17	W00127	ME#2a library gene
13	38	79.2	73	21	G02472	Human secreted pro
14	38	79.2	246	14	R40924	Protein able to bi
15	38	79.2	470	14	R34476	Encoded by Hepatit
16	38	79.2	470	14	R34477	Encoded by Hepatit
17	38	79.2	470	14	R34478	Encoded by Hepatit
18	36	75.0	218	20	V73909	Human prostate tum
19	36	75.0	244	21	V65525	Human gene 72-enco
20	36	75.0	267	19	W41964	Flea serine protea
21	36	75.0	267	22	B50625	Novel serine protea
22	36	75.0	1330	22	B65630	Human zneul BGF-11
23	35	72.9	31	20	W88384	ME#2a and ME#2b 11
24	35	72.9	38	17	R91412	Human zneul parlia
25	35	72.9	73	20	W88389	Human zneul parlia
26	35	72.9	89	21	G01562	Human secreted pro
27	35	72.9	115	21	B41718	Human ORFX ORF1482
28	35	72.9	153	21	B41638	Human ORFX ORF1402
29	35	72.9	158	20	W88388	Human zneul parlia
30	35	72.9	169	20	W88390	Human zneul parlia
31	35	72.9	187	21	V72094	Human serine prote
32	35	72.9	224	20	V59870	Human normal uteru
33	35	72.9	238	21	B11695	Mouse serine prote
34	35	72.9	247	21	V52139	Human TANGO 125b (
35	35	72.9	254	20	W88382	Human neuro-growth
36	35	72.9	265	21	B42204	Human ORFX ORF1968
37	35	72.9	273	20	V41769	Human PRO213-1 pro
38	35	72.9	273	20	V41770	Human PRO1330 prot
39	35	72.9	273	20	V41771	Human PRO1449 prot
40	35	72.9	273	20	W88381	Human neuro-growth
41	35	72.9	273	21	B44325	Human PRO213-1 pro
42	35	72.9	273	21	B44326	Human PRO1330 prot
43	35	72.9	273	21	B44327	Human PRO1449 prot
44	35	72.9	273	21	B18673	Amino acid sequenc
45	35	72.9	273	21	B18674	Amino acid sequenc

## ALIGNMENTS

RESULT	ID	Y79113	standard; Peptide; 8 AA.
1	Y79113		
AC	Y79113:		
XX	05-JUN-2000	(first entry)	
XX	Peptide antagonist of zonulin.		
DE			
KW	Zonulin: antagonist; zonula occludens toxin receptor;		
KW	blood-brain barrier; antiinflammatory; cerebroprotective;		
KW	neuroprotective; dermatological; antitumor; antiviral;		
KW	antibacterial; cytostatic; anti-HIV; vulnary; antiallergic;		
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;		
KW	gastrointestinal inflammation; therapy.		
OS	Synthetic.		
PN	WO200007609-A1.		
PD	17-FEB-2000.		
XX			
XX	28-JUL-1999:	99MO-US16683.	
XX	03-AUG-1998:	98US-0127815.	
XX	(UYMA-) UNIV MARYLAND BALTIMORE.		
XX	Fasano A:		
XX	WPI: 2000-205565/18.		
XX	New peptide antagonist of zonulin useful as antiinflammatory agent for		

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1; Page 43; 69pp; English.

XX This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see W79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 XX  
 SQ Sequence 8 AA:

Query Match 100.0%; Score 48; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 1 |||||||  
 Db 1 ggcvcvpg 8

RESULT 2  
 ID W79125 standard; Peptide; 8 AA.  
 AC W79125;

DT 05-JUN-2000 (first entry)  
 XX

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

XX Synthetic.

OS W0200007609-A1.

PN 17-FEB-2000.

PD 28-JUL-1999; 99WO-US16683.

XX 03-AUG-1998; 98US-0127815.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

PI Fasano A;  
 XX WPI; 2000-205565/18.  
 DR

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1; Page 47; 69pp; English.

XX This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see W79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 XX  
 SQ Sequence 8 AA:

Query Match 85.4%; Score 41; DB 21; Length 8;  
 Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 1 |||||||  
 Db 1 ggcvcvpg 8

RESULT 3  
 ID W79114 standard; Peptide; 8 AA.  
 AC W79114;

DT 05-JUN-2000 (first entry)  
 XX

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antitumor; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

XX Synthetic.

OS W0200007609-A1.

PN 17-FEB-2000.

PD 28-JUL-1999; 99WO-US16683.



XX 03-AUG-1998: 98US-0127815.  
XX (UYMA-) UNIV MARYLAND BALTIMORE.  
XX Fasano A;  
XX WPI: 2000-205565/18.  
XX  
XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
XX treating cerebral ischemia, stroke, cerebral edema, gastritis,  
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
XX Claim 1: Page 44; 69pp; English.  
XX  
XX This present sequence is that of a peptide antagonist of zonulin  
XX (2), one of 25 such peptides (see Y79105-29) of the invention,  
XX which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
XX physiologically modulate the opening of mammalian tight junctions  
XX (TJ). The peptide antagonists are based on a common motif of ZOT  
XX and human zonulin, which is believed to be critical for receptor  
XX binding. They can be prepared by chemical synthesis or by use of  
XX recombinant DNA techniques. The peptide antagonists are used as an  
XX antiinflammatory agents in the treatment of gastrointestinal  
XX inflammation, where they bind to the ZOT receptor in the intestine  
XX and yet does not physiologically modulate the opening of TJ in the  
XX intestine. Gastrointestinal inflammation conditions give rise to  
XX increased intestinal permeability and the peptide is useful for  
XX treating intestinal conditions that cause protein losing enteropathy  
XX caused by infection, e.g. Clostridium difficile infection,  
XX enterocolitis, shigellosis, viral gastroenteritis, parasite  
XX infestation, bacterial overgrowth, whipple's disease, diseases with  
XX mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
XX collagenous colitis, inflammatory bowel disease, diseases marked by  
XX lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
XX sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
XX correction of congenital heart disease with Fontan's operation,  
XX mucosal diseases without ulceration, e.g. Menetrier's disease,  
XX coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
XX e.g. systemic lupus erythematosus or food allergies, primarily to  
XX milk.  
XX  
XX Sequence 8 AA:  
XX  
XX  
XX Query Match 83.3%; Score 40; DB 21; Length 8;  
XX Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 GRGCVOPG 8  
XX | | | | | | | |  
XX Db 1 grgcvgdg 8  
XX  
XX  
XX RESULT 4  
XX Y79105  
XX ID Y79105 standard; Peptide: 8 AA.  
XX  
XX Y79105:  
XX  
XX 05-JUN-2000 (first entry)  
XX  
XX Peptide antagonist of zonulin.  
XX  
XX Zonulin; antagonist; zonula occludens toxin receptor;  
XX blood-brain barrier; antiinflammatory; cerebroprotective;  
XX neuroprotective; dermatological; antitumor; antiviral;  
XX antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
XX hypotensive; immunosuppressive; antiparasitic; vasotropic;  
XX gastrointestinal inflammation; therapy.  
XX  
XX Synthetic.  
XX  
XX

PN W0200007609-A1.  
XX  
XX 17-FEB-2000.  
XX  
XX 28-JUL-1999: 99MO-US16683.  
XX  
XX 03-AUG-1998: 98US-0127815.  
XX  
XX (UYMA-) UNIV MARYLAND BALTIMORE.  
XX Fasano A;  
XX WPI: 2000-205565/18.  
XX  
XX New peptide antagonist of zonulin useful as antiinflammatory agent for  
XX treating cerebral ischemia, stroke, cerebral edema, gastritis,  
XX shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
XX Claim 1: Page 41; 69pp; English.  
XX  
XX This present sequence is that of a peptide antagonist of zonulin  
XX (2), one of 25 such peptides (see Y79105-29) of the invention,  
XX which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
XX physiologically modulate the opening of mammalian tight junctions  
XX (TJ). The peptide antagonists are based on a common motif of ZOT  
XX and human zonulin, which is believed to be critical for receptor  
XX binding. They can be prepared by chemical synthesis or by use of  
XX recombinant DNA techniques. The peptide antagonists are used as an  
XX antiinflammatory agents in the treatment of gastrointestinal  
XX inflammation, where they bind to the ZOT receptor in the intestine  
XX and yet does not physiologically modulate the opening of TJ in the  
XX intestine. Gastrointestinal inflammation conditions give rise to  
XX increased intestinal permeability and the peptide is useful for  
XX treating intestinal conditions that cause protein losing enteropathy  
XX caused by infection, e.g. Clostridium difficile infection,  
XX enterocolitis, shigellosis, viral gastroenteritis, parasite  
XX infestation, bacterial overgrowth, whipple's disease, diseases with  
XX mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
XX collagenous colitis, inflammatory bowel disease, diseases marked by  
XX lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
XX sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
XX correction of congenital heart disease with Fontan's operation,  
XX mucosal diseases without ulceration, e.g. Menetrier's disease,  
XX coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
XX e.g. systemic lupus erythematosus or food allergies, primarily to  
XX milk.  
XX  
XX Sequence 8 AA:  
XX  
XX  
XX Query Match 81.2%; Score 39; DB 21; Length 8;  
XX Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 1 GRGCVOPG 8  
XX | | | | | | | |  
XX Db 1 grgcvgpg 8  
XX  
XX  
XX RESULT 5  
XX Y79109  
XX ID Y79109 standard; Peptide: 8 AA.  
XX  
XX Y79109:  
XX  
XX 05-JUN-2000 (first entry)  
XX  
XX Peptide antagonist of zonulin.  
XX  
XX Zonulin; antagonist; zonula occludens toxin receptor;  
XX blood-brain barrier; antiinflammatory; cerebroprotective;  
XX neuroprotective; dermatological; antitumor; antiviral;  
XX antibacterial; cytostatic; anti-HIV; vulnerary; antiallergic;  
XX

XX Zonulin; antagonist; zonula occludens toxin receptor;  
KM blood-brain barrier; antiinflammatory; cerebroprotective;  
KM neuroprotective; dermatological; anticancer; antiviral;  
KM antibacterial; cytostatic; anti-HIV; vulnuary; antiallergic;  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
XX gastrointestinal inflammation; therapy..  
OS Synthetic.  
XX  
XX WO200007609-A1.  
PN  
PD 17-FEB-2000.  
XX  
PF 28-JUL-1999; 99WO-US16683.  
XX  
PR 03-AUG-1998; 98US-0127815.  
XX  
PA (UYMA-) UNIV MARYLAND BALTIMORE.  
XX  
PI Pasano A:  
XX  
DR WPI: 2000-205565/18.  
XX  
PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -  
XX  
PS Claim 1; Page 44; 69pp: English.

This present sequence is that of a peptide antagonist of zonulin (Z), one of 25 such peptides (see Y79105-29) of the invention, which bind to a zonula occludens toxin (ZOT) receptor, yet do not physiologically modulate the opening of mammalian tight junctions (TJ). The peptide antagonists are based on a common motif of ZOT and human zonulins, which is believed to be critical for receptor binding. They can be prepared by chemical synthesis or by use of recombinant DNA techniques. The peptide antagonists are used as an antiinflammatory agents in the treatment of gastrointestinal inflammation, where they bind to the ZOT receptor in the intestine and yet does not physiologically modulate the opening of TJ in the intestine. Gastrointestinal inflammation conditions give rise to increased intestinal permeability and the peptide is useful for treating intestinal conditions that cause protein losing enteropathy caused by infection, e.g. Clostridium difficile infection, CC enterocolitis, shigellosis, viral gastroenteritis, parasite infestation, bacterial overgrowth, whipple's disease, diseases with mucosal erosion or ulcerations, e.g. gastritis, gastric cancer, CC collagenous colitis, inflammatory bowel disease, diseases marked by lymphatic obstruction, e.g. congenital intestinal lymphangiectasia, sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical correction of congenital heart disease with Fontan's operation, CC mucosal diseases without ulceration, e.g. Menetrier's disease, CC coeliac disease, eosinophilic gastroenteritis, and immune diseases, e.g. systemic lupus erythematosus or food allergies, primarily to milk.

Sequence 8 AA;

Query Match 79.2%; Score 38; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVPG 8  
| | | | |  
Db 1 grglvpg 8

RESULT 7  
R91407  
ID R91407 standard; Peptide; 38 AA.  
X

AC R91407;  
 XX  
 PD 10-OCT-1996 (first entry)  
 DT XX  
 DE ME#2a and ME#2b library derived peptide, J5-2A8.  
 XX  
 KW Cis-element; ligand; highly specific DNA binding domain; HSDB; human;  
 KW H2-kappa-B; regulatory domain; murine; homologue; IL-6; Interleukin;  
 KW HLA class I gene; HLA-B\*27; HLA-A\*2; acute phase protein response;  
 KW syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;  
 KW transcription.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9606188-A1.  
 PD 29-FEB-1996.  
 XX  
 PF 17-AUG-1995; 95WO-US10523.  
 XX  
 PR 15-AUG-1995; 95US-0515190.  
 PR 18-AUG-1994; 94US-0292902.  
 XX  
 PA (CYTO-) CYTOGEN CORP.  
 XX  
 PI Fowlkes DM, Rodwell JD;  
 XX  
 DR WPI; 1996-151391/15.  
 XX  
 PT Synthetic random nucleotide sequences encoding ligand binding  
 PT domains - identified by screening library of vectors or peptide(s)  
 PT and useful for gene therapy and diagnosis  
 XX  
 PS Claim 70; Page 197; 224pp: English.  
 XX  
 CC The sequences given in R91378-414 represent peptides that bind to  
 CC highly specific DNA binding domains (HSDB)'s. These sequences were  
 CC tested for binding to the H2-kappa-B oligonucleotide which contains  
 CC the NF-kappa-B binding site, and comprises the sequence given in T13579.  
 CC These peptides may be used in a composition for diagnosis and gene  
 CC therapy and for modifying the transcription or activity of a gene.  
 CC  
 XX  
 SQ Sequence 38 AA;

Query Match 79.2%; Score 38; DB 17; Length 38;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 | | | | |  
 Db 14 grgcfnpg 21

RESULT 8  
 R91405  
 ID R91406 standard; Peptide: 38 AA.  
 XX  
 AC R91406;  
 XX  
 DT 10-OCT-1996 (first entry)  
 DT XX  
 DE ME#2a and ME#2b library derived peptide, J5-2A4.  
 XX  
 KW Cis-element; ligand; highly specific DNA binding domain; HSDB; human;  
 KW H2-kappa-B; regulatory domain; murine; homologue; IL-6; Interleukin;  
 KW HLA class I gene; HLA-B\*27; HLA-A\*2; acute phase protein response;  
 KW syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;  
 KW transcription.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9606188-A1.  
 PD 29-FEB-1996.  
 XX  
 PF 17-AUG-1995; 95WO-US10523.  
 XX  
 PR 15-AUG-1995; 95US-0515190.  
 PR 18-AUG-1994; 94US-0292902.  
 XX  
 PA (CYTO-) CYTOGEN CORP.  
 XX  
 PI Fowlkes DM, Rodwell JD;  
 XX  
 DR WPI; 1996-151391/15.  
 XX  
 PT Synthetic random nucleotide sequences encoding ligand binding

XX  
 PD 29-FEB-1996.  
 XX  
 PF 17-AUG-1995; 95WO-US10523.  
 XX  
 PR 15-AUG-1995; 95US-0515190.  
 PR 18-AUG-1994; 94US-0292902.  
 XX  
 PA (CYTO-) CYTOGEN CORP.  
 XX  
 PI Fowlkes DM, Rodwell JD;  
 XX  
 DR WPI; 1996-151391/15.  
 XX  
 PT Synthetic random nucleotide sequences encoding ligand binding  
 PT domains - identified by screening library of vectors or peptide(s)  
 PT and useful for gene therapy and diagnosis  
 XX  
 PS Claim 70; Page 197; 224pp: English.  
 XX  
 CC The sequences given in R91378-414 represent peptides that bind to  
 CC highly specific DNA binding domains (HSDB)'s. These sequences were  
 CC tested for binding to the H2-kappa-B oligonucleotide which contains  
 CC the NF-kappa-B binding site, and comprises the sequence given in T13579.  
 CC These peptides may be used in a composition for diagnosis and gene  
 CC therapy and for modifying the transcription or activity of a gene.  
 CC  
 XX  
 SQ Sequence 38 AA;

Query Match 79.2%; Score 38; DB 17; Length 38;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 | | | | |  
 Db 14 grgcfnpg 21

RESULT 9  
 R91405  
 ID R91405 standard; Peptide: 38 AA.  
 XX  
 AC R91405;  
 XX  
 DT 10-OCT-1996 (first entry)  
 DT XX  
 DE ME#2a and ME#2b library derived peptide, J5-1A10.  
 XX  
 KW Cis-element; ligand; highly specific DNA binding domain; HSDB; human;  
 KW H2-kappa-B; regulatory domain; murine; homologue; IL-6; Interleukin;  
 KW HLA class I gene; HLA-B\*27; HLA-A\*2; acute phase protein response;  
 KW syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;  
 KW transcription.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9606188-A1.  
 PD 29-FEB-1996.  
 XX  
 PF 17-AUG-1995; 95WO-US10523.  
 XX  
 PR 15-AUG-1995; 95US-0515190.  
 PR 18-AUG-1994; 94US-0292902.  
 XX  
 PA (CYTO-) CYTOGEN CORP.  
 XX  
 PI Fowlkes DM, Rodwell JD;  
 XX  
 DR WPI; 1996-151391/15.  
 XX  
 PT Synthetic random nucleotide sequences encoding ligand binding



```

DT 15-OCT-1996 (first entry)
XX
DE ME#2a library generic peptide sequence.
XX
KM C1s-element; ligand; highly specific DNA binding domain; HSDb; human;
KM NF-kappa-B; regulatory domain; murine; homologue; IL-6; Interleukin;
KM HLA class I gene; HLA-B*27; HLA-J; HLA-A*2; acute phase protein response;
KM syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;
KM transcription.
XX
OS Synthetic.
XX
FH Key
FT MISC-difference 3 Location/Qualifiers
FT MISC-difference 4..13 /label= Ser, Arg
FT MISC-difference 31..40 /note= "Any amino acid"
FT MISC-difference 31..40 /note= "Any amino acid"
XX
PN WO9606188-A1.
XX
PD 29-FEB-1996.
XX
PF 17-AUG-1995; 95WO-US10523.
XX
PR 15-AUG-1995; 95US-0515190.
PR 18-AUG-1994; 94US-0292902.
XX
PA (CYTO-) CYTOGEN CORP.
XX
PI Fowles DM, Rodwell JD;
XX
DR WPI; 1996-151391/15.
XX
PT Synthetic random nucleotide sequences encoding ligand binding
PT domains - identified by screening library of vectors or peptide(s)
PT and useful for gene therapy and diagnosis
XX
PS Disclosure; Page 17; 224pp; English.
XX
CC The sequences given in T13619 and T13622-23 represent oligonucleotides
CC which were used in the generation of the ME#2a random peptide library.
CC T13619 encodes a region of the wildtype peptide given in W00127, which
CC is replaced by the random peptides in the finished library. The
CC sequences of amino acids generated in the random library are
CC unpredictable and substantially random in sequence. All 20 naturally
CC occurring amino acids are encoded by the variable coding region.
CC Invariant nucleotides are positioned at particular sites in these
CC oligonucleotide sequences to aid the assembly and cloning of the synthesised
CC oligonucleotides. At the 5' terminus of each set of the oligonucleotides,
CC the invariant nucleotides encode for efficient restriction endonuclease
CC cleavage sites. The invariant nucleotides at the 5' terminus are chosen
CC to encode pairs of sites for by restriction enzymes which can function in
CC the same buffer conditions, are commercially available at high specific
CC activity, are not complementary to each other to prevent self-ligation
CC of the synthesised double stranded oligonucleotides and which require 6
CC or 8 nucleotides for a cleavage recognition site in order to lower the
CC frequency of cleaving within the inserted double stranded synthesised
CC oligonucleotide sequences. The 3' terminal invariant nucleotide
CC positions are complementary pairs of 6, 9 or 12 nucleotides to aid in
CC annealing two synthesised single stranded sets of nucleotides together
CC and conversion to double stranded DNA. See also R91378-414.
XX
SQ Sequence 42 AA;

```

```

Query Match 79.2%; Score 38; DB 17; Length 42;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GRCGVOPG 8
|||||

```

```

DB 16 grgcfnpg 23
RESULT 13
ID G02472
ID G02472 standard; Protein; 73 AA.
XX
AC G02472;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 6553.
XX
KM Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Maline Edwards J, Duclet A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
DR N-PSDB; C02478.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 6553; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 73 AA;

```

```

Query Match 79.2%; Score 38; DB 21; Length 73;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GRCGVOPG 8
|||||
DB 43 grgcnepg 50
RESULT 14
ID R40924
ID R40924 standard; Protein; 246 AA.
XX
AC R40924;
XX
DT 17-FEB-1994 (first entry)
XX

```

DE Protein able to bind to HIV-1 tat protein.  
 XX  
 KW Antibodies; antigen binding proteins; library; HIV;  
 KW Human Immunodeficiency Virus.  
 XX  
 OS Synthetic.  
 XX EP557897-A.  
 PN  
 XX 01-SEP-1993.  
 PD  
 XX 19-FEB-1993; 93EP-0102609.  
 PE  
 XX 28-FEB-1992; 92US-0843125.  
 PR  
 XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
 PA  
 XX  
 PI Dillon PJ, Rosen CA.  
 XX  
 DR WPI; 1993-274375/35.  
 DR N-PSDB; Q48605.  
 XX  
 PT Antigen-binding proteins and corresp. synthetic genes - are  
 PT constructed synthetically and used to make antibody library, in  
 PT disease diagnosis, etc.  
 XX  
 PS Claim 13; Page 22; 40pp; English.  
 XX  
 CC The synthetic gene encodes a protein corresponding to an antibody  
 CC capable of binding to a specific antigen, in this case the HIV-1 tat  
 CC protein. Many synthetic genes are synthesised, each containing a  
 CC predetermined nucleotide region encoding the framework regions of  
 CC the heavy and light chains of antibody and undetermined nucleotide  
 CC regions which are random sequences. The genes are then used in the  
 CC construction of vectors which are subsequently used to transform  
 CC microbes. The proteins thus produced are screened for binding  
 CC activity to the specific antigen.  
 CC  
 SQ Sequence 246 AA;  
 XX

Query Match 79.2%; Score 38; DB 14; Length 246;  
 Best Local Similarity 87.3%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRGCVQPG 8  
 ||| |||  
 Db 8 grglvqpg 15

RESULT 15  
 R34476  
 ID R34476 standard; Protein; 470 AA.  
 XX  
 AC R34476;  
 XX  
 DT 30-JUL-1993 (first entry)  
 XX  
 DE Encoded by Hepatitis C virus clone JK4-A.  
 XX  
 KW HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;  
 KW polymerase chain reaction; diagnostic method.  
 XX  
 OS Hepatitis C virus.  
 OS  
 PN JF05068562-A.  
 PD  
 XX 23-MAR-1993.  
 PD  
 XX 30-MAY-1991; 91JP-0153736.  
 PF  
 XX 30-MAY-1991; 91JP-0153736.  
 PR  
 XX

PA (SANW ) SANWA KAGAKU KENKYUSHO CO.  
 XX  
 DR WPI; 1993-130638/16.  
 DR N-PSDB; Q40434.  
 XX  
 PT DNA and cDNA of hepatitis C virus - useful as probes for  
 PT diagnosing HCV infection  
 XX  
 PS Claim 4; Page 32-34; 44pp; Japanese.  
 XX  
 CC cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B  
 CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR  
 CC amplification, including clone JK4-A. Primer/probes derived from the  
 CC sequences of these clones can be used in diagnostic assays for HCV.  
 CC See Q40425-Q40439.  
 CC  
 SQ Sequence 470 AA;  
 XX

Query Match 79.2%; Score 38; DB 14; Length 470;  
 Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 GRGCVQPG 8  
 ||| |||  
 Db 73 gracaqpg 80

Search completed: June 13, 2001, 14:14:34  
 Job time: 375 sec



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:41 ; Search time 130.61 Seconds  
(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-9

Perfect score: 48

Sequence: 1 GRGCVPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_67:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Score	Match length	ID	Description
1	85.4	2240	2 T37057	probable multi-dom
2	83.3	411	2 S41648	translation elonga
3	79.2	267	2 S40006	trypsin (EC 3.4.21
4	79.2	275	2 S40005	trypsin (EC 3.4.21
5	79.2	782	2 S18032	genome polypeptide
6	77.1	277	2 S35340	trypsin (EC 3.4.21
7	77.1	534	2 C70548	probable regulator
8	77.1	728	1 TVHUSK	transforming prote
9	75.0	471	2 H83460	probable carboxydr
10	72.9	145	2 H69051	heterodisulfide re
11	72.9	272	2 JC4170	trypsin-like prote
12	72.9	433	2 T75566	probable benzate
13	72.9	525	2 T24647	hypothetical prote
14	72.9	593	2 S49525	glycoprotein G - s
15	72.9	706	2 D83432	type III secretory
16	70.8	231	2 S31778	trypsin (EC 3.4.21
17	70.8	242	2 S31775	trypsin (EC 3.4.21
18	70.8	242	2 S31776	trypsin (EC 3.4.21
19	70.8	250	2 T01779	trypsin (EC 3.4.21
20	70.8	303	2 S31384	trypsin (EC 3.4.21
21	70.8	317	2 T13598	trypsin homolog -
22	70.8	357	1 H70566	hypothetical prote
23	70.8	478	1 K1BE8R	thymidine kinase (
24	70.8	478	1 A72721	probable glycerol
25	70.8	480	1 A45339	outer capsid prote
26	70.8	680	1 PNO510	integrin beta-3 ch
27	70.8	723	2 PNO509	integrin beta-3 ch
28	70.8	842	2 T04555	hypothetical prote
29	70.8	1159	2 T02866	hypothetical prote

30	33	68.8	92	2 D37057	epithelial cell gl
31	33	68.8	162	2 C72699	hypothetical prote
32	33	68.8	196	2 T38022	hypothetical prote
33	33	68.8	226	2 T20645	hypothetical prote
34	33	68.8	237	1 TRCY1	trypsin (EC 3.4.21
35	33	68.8	259	1 TRSMG	trypsin (EC 3.4.21
36	33	68.8	266	2 S54146	trypsin (EC 3.4.21
37	33	68.8	273	2 A47246	trypsin (EC 3.4.21
38	33	68.8	275	2 A35863	trypsin (EC 3.4.21
39	33	68.8	275	2 B35863	trypsin (EC 3.4.21
40	33	68.8	275	2 C35863	trypsin (EC 3.4.21
41	33	68.8	275	2 A32410	trypsin (EC 3.4.21
42	33	68.8	276	2 A38654	mast cell proteina
43	33	68.8	404	1 VHYNHS	nucleoprotein - he
44	33	68.8	404	1 J01531	nucleoprotein - he
45	33	68.8	404	2 S34562	nucleocapsid prote

## ALIGNMENTS

RESULT 1  
T37057  
probable multi-domain beta keto-acyl synthase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 03-Dec-1999 #sequence=revision 03-Dec-1999 #text\_change 28-Jul-2000  
C:Accession: T37057  
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: 221620  
A:Accession: T37057  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2240 <SEP>  
A:Cross-references: EMBL:AL109747; PDB:GAB52354.1; GSPDB:GN00070; SCODEB:SCJ21.08  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCODEB:SCJ21.08  
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology  
F:24-435/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <AS>

Query Match 85.4% Score 41; DB 2; Length 2240;  
Best Local Similarity 87.5% Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GRGCVPG 8  
DB 10 GRGCVLP 17

RESULT 2  
S41648  
translation elongation factor eEF-1 gamma - Trypanosoma cruzi  
C:Species: Trypanosoma cruzi  
C>Date: 20-May-1994 #sequence=revision 26-May-1995 #text\_change 26-Aug-1999  
C:Accession: S41648  
R:Billaut-Mulot, O.; Pommer, V.; Schoeneck, R.; Plumas-Marty, B.; Taib, A.; Loyens, N.  
Nucleic Acids Res. 21, 3901, 1993  
A>Title: Nucleotide sequence of a Trypanosoma cruzi cDNA encoding a protein homologous  
A:Reference number: S41648; MUID:93376525  
A:Accession: S41648  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-411 <BI>  
A:Cross-references: EMBL:L17307; NID:g349148; PID:g349149  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1993  
C:Superfamily: translation elongation factor eEF-1 gamma chain

Query Match 83.3% Score 40; DB 2; Length 411;  
Best Local Similarity 75.0% Pred. No. 9.1;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVOPG 8  
Db 222 GRGCAOPG 229

## RESULT 3

trypsin (EC 3.4.21.4) precursor - African malaria mosquito  
S40006  
C:Species: Anopheles gambiae (African malaria mosquito)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 22-Jun-1999  
C:Accession: S40006  
R:Mueller, H.  
Submitted to the EMBL Data Library, June 1993  
A:Reference number: S40003  
A:Accession: S40006  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-267 <MUE>  
A:Cross-references: EMBL:Z22930; NID:G410323; PIDN:CAA80516.1; PID:G410327  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F:42-261/Domain: trypsin homology <TRY>  
F:82/127/222/Active site: His, Asp, Ser #status predicted

Query Match 79.2%; Score 38; DB 2; Length 267;  
Best Local Similarity 75.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GRGCVOPG 8  
Db 239 GAGCAOPG 246

## RESULT 4

trypsin (EC 3.4.21.4) precursor - African malaria mosquito  
S40005  
C:Species: Anopheles gambiae (African malaria mosquito)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 22-Jun-1999  
C:Accession: S40005  
R:Mueller, H.  
Submitted to the EMBL Data Library, June 1993  
A:Reference number: S40003  
A:Accession: S40005  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-275 <MUE>  
A:Cross-references: EMBL:Z22930; NID:G410323; PIDN:CAA80515.1; PID:G410326  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F:49-269/Domain: trypsin homology <TRY>  
F:89/134/230/Active site: His, Asp, Ser #status predicted

Query Match 79.2%; Score 38; DB 2; Length 275;  
Best Local Similarity 75.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GRGCVOPG 8  
Db 247 GAGCAOPG 254

## RESULT 5

genome polypeptide - hepatitis C virus (isolate JK4) (fragment)  
S18032  
N:Conting: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein  
C:Species: hepatitis C virus  
A:Variety: isolate JK4  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000  
C:Accession: S18032  
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991  
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus  
A:Reference number: S18029

A:Accession: S18032  
A:Molecule type: genomic RNA  
A:Residues: 1-782 <HON>  
A:Cross-references: EMBL:X61594  
A:Experimental source: isolate JK4  
C:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural  
F:1-191/Product: core protein #status predicted <MAT1>  
F:192-383/Product: envelope protein 1 #status predicted <MAT2>  
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>  
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 79.2%; Score 38; DB 2; Length 782;  
Best Local Similarity 75.0%; Pred. No. 35;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GRGCVOPG 8  
Db 73 GRACAPG 80

## RESULT 6

trypsin (EC 3.4.21.4) 2 precursor - African malaria mosquito  
S35340  
C:Species: Anopheles gambiae (African malaria mosquito)  
C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 22-Jun-1999  
C:Accession: S35340; S40008; S35413  
R:Mueller, H.M.; Crampson, J.M.; della Torre, A.; Sindén, R.; Crisanti, A.  
EMBO J. 12, 2891-2900, 1993  
A:Title: Members of a trypsin gene family in Anopheles gambiae are induced in the gut  
A:Reference number: S35339; MID:93327779

A:Accession: S35340  
A:Molecule type: DNA  
A:Residues: 1-277 <MUE>  
A:Cross-references: EMBL:Z18890; NID:G312023; PIDN:CAA79328.1; PID:G312024  
A:Experimental source: strain Snakoko  
R:Mueller, H.  
Submitted to the EMBL Data Library, June 1993  
A:Reference number: S40003  
A:Accession: S40008  
A:Molecule type: DNA  
A:Residues: 1-277 <MUE>  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase; zymogen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-56/Domain: activation peptide #status predicted <APT>  
F:51-277/Product: trypsin 2 #status predicted <MAT>  
F:51-271/Domain: trypsin homology <TRY>  
F:91/136/232/Active site: His, Asp, Ser #status predicted

Query Match 77.1%; Score 37; DB 2; Length 277;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GRGCVOPG 8  
Db 249 GYCAOPG 256

## RESULT 7

probable regulatory protein - Mycobacterium tuberculosis (strain H37RV)  
C70548  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: C70548  
R:Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: C70548  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-534 <COL>  
 A:Cross-references: GB:295558; GB:AL123456; NID:g3261781; PID:CAB08963.1; PID:g2114014  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0552

Query Match 77.1%; Score 37; DB 2; Length 534;  
 Best Local Similarity 75.0%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCGVOP 8  
 1 11111  
 DB 55 GAGCVMFG 62

RESULT 8  
 TVHUSK  
 transforming protein ski - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 18-Jun-1999  
 C:Accession: S06053  
 R:Nomura, N.; Sasamoto, S.; Ishii, S.; Date, T.; Matsui, M.; Ishizaki, R.  
 Nucleic Acids Res. 17, 5489-5500, 1989  
 A:Title: Isolation of human cDNA clones of ski and the ski-related gene, sno.  
 A:Reference number: S06052; MUID:89345144  
 A:Accession: S06053  
 A:Molecule type: mRNA  
 A:Residues: 1-728 <NON>  
 A:Cross-references: GB:X15218; NID:g36483; PIDN:CAA33288.1; PID:g36484  
 C:Comment: Transforming proteins ski may function as a transcription factor regulating m  
 C:Genetics:  
 A:Gene: GDB:SKI  
 A:Cross-references: GDB:119595; OMIM:164780  
 A:Map position: 1q22-1q24  
 C:Superfamily: ski transforming protein  
 C:Keywords: proto-oncogene, tandem repeat; transforming protein  
 F:61-93/Region: proline-rich  
 F:494-509/Region: serine-rich  
 F:558-582,583-607,608-633/Region: 25-residue repeats

Query Match 77.1%; Score 37; DB 1; Length 728;  
 Best Local Similarity 85.7%; Pred. No. 50;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCGVOP 7  
 1 11111  
 DB 7 GRCGRFP 13

RESULT 9  
 HB3460  
 probable carbohydrate kinase PA1487 [Imported] - *Pseudomonas aeruginosa* (strain PA01)  
 C:Species: *Pseudomonas aeruginosa*  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: HB3460  
 R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: HB3460  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-471 <STO>  
 A:Cross-references: GB:AE004577; GB:AE004091; NID:g9947430; PIDN:AA04876.1; GSPD:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA1487  
 C:Superfamily: xylokinease

Query Match 75.0%; Score 36; DB 2; Length 471;  
 Best Local Similarity 75.0%; Pred. No. 52;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCGVOP 8  
 1 11111  
 DB 232 GRCGRFP 239

RESULT 10  
 HB9051  
 heterodisulfide reductase Hrd related protein - *Methanobacterium thermoautotrophicum*  
 C:Species: *Methanobacterium thermoautotrophicum*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: H69051  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T  
 .; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani,  
 K.L.S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: fu  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: H69051  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-145 <MTB>  
 A:Cross-references: GB:AE00803; GB:AE000666; NID:g2621179; PIDN:AAB84645.1; PID:g262  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH139

Query Match 72.9%; Score 35; DB 2; Length 145;  
 Best Local Similarity 85.7%; Pred. No. 29;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCGVOP 7  
 1 11111  
 DB 100 GCGGVOP 106

RESULT 11  
 JC4170  
 trypsin-like proteinase (EC 3.4.21.-) precursor - *Saccharopolyspora erythraea*  
 C:Species: *Saccharopolyspora erythraea*  
 C:Date: 27-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Jun-2000  
 C:Accession: JC4170; PC4045  
 R:Nagamine-Watsuka, Y.; Norioka, S.; Sakiyama, F.  
 J. Biochem. 118, 338-346, 1995  
 A:Title: Molecular cloning, nucleotide sequence, and expression of the gene encoding  
 A:Reference number: JC4170; MUID:96064153  
 A:Note: Species designated as *Streptomyces erythraeus*  
 A:Accession: JC4170  
 A:Molecule type: DNA  
 A:Residues: 1-212 <NAG>  
 A:Cross-references: DDBJ:D30760; NID:g493712; PIDN:BA06423.1; PID:9871932  
 A:Accession: PC4045  
 A:Molecule type: protein  
 A:Residues: 42-51 <NAG>  
 C:Comment: This protein is a trypsin-like serine proteinase but is distinct from tryp  
 sin autolysis.  
 C:Genetics:  
 A:Gene: set  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase  
 F:1-30/Domain: signal sequence #status predicted <SIG>

F:31-41/Domain: propeptide #status predicted <PRO>  
F:42-772/Product: trypsin-like proteinase #status predicted <MAT>  
F:43-260/Domain: trypsin homology <TRY>

Query Match 72.9%; Score 35; DB 2; Length 272;  
Best Local Similarity 62.5%; Pred. No. 49;  
Matches: 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
DB 238 GEGCARG 245

## RESULT 12

F75566  
probable benzate membrane transport protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75566

R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: F75566

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <WHI>

A:Cross-references: GB:AE001868; GB:AE000513; NID:g6457707; PIDN:AAF09646.1; PID:g645771

A:Experimental source: strain R1

A:Gene: DR0056

A:Map position: 1

Query Match 72.9%; Score 35; DB 2; Length 433;  
Best Local Similarity 75.0%; Pred. No. 73;  
Matches: 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
DB 383 GRGAOPG 390

## RESULT 13

T24647

hypothetical protein T07C5.1a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T24647

R.McMurray, A.

submitted to the EMBL Data Library, July 1995

A:Reference number: Z19917

A:Accession: T24647

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-525 <WIL>

A:Cross-references: EMBL:Z50006; PIDN:CAA90296.1; GSPDB:GN00028; CESP:T07C5.1a

A:Experimental source: clone T07C5

A:Gene: CESP:T07C5.1a

A:Map position: X

A:Intons: 39/3; 106/1; 139/2; 231/3; 290/3; 344/1; 371/1; 420/3; 444/2

C:Superfamily: g1ucuronosyltransferase

Query Match 72.9%; Score 35; DB 2; Length 525;  
Best Local Similarity 71.4%; Pred. No. 85;  
Matches: 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGCVOP 7

DB 158 GOGCVBP 164

## RESULT 14

S49525

glycoprotein G - simian herpesvirus B

N:Alternate names: US4 protein homolog

C:Species: simian herpesvirus B

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999

C:Accession: S49525

R.Slomka, M.J.; Brown, D.W.

submitted to the EMBL Data Library, October 1994

A:Description: Complete nucleotide sequence of simian herpes B virus glycoprotein G 9

A:Reference number: S49525

A:Accession: S49525

A:Molecule type: DNA

A:Residues: 1-593 <SLO>

A:Cross-references: EMBL:Z46268; NID:g560495; PIDN:CAA86431.1; PID:g560496

A:Experimental source: isolate Cyno 2

C:Keywords: glycoprotein

Query Match 72.9%; Score 35; DB 2; Length 593;  
Best Local Similarity 75.0%; Pred. No. 95;  
Matches: 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
DB 248 GRGCVSPG 255

## RESULT 15

D83432

type III secretory apparatus protein Pcrd PA1703 [imported] - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: D83432

R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: AB2950; MUID:20437337

A:Accession: D83432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-706 <STO>

A:Cross-references: GB:AE004597; GB:AE004091; NID:g9947671; PIDN:ANG05092.1; GSPDB:GN

A:Experimental source: strain PA01

C:Gene: pcrd; PA1703

C:Superfamily: regulatory protein Icrd

Query Match 72.9%; Score 35; DB 2; Length 706;  
Best Local Similarity 71.4%; Pred. No. 11e+02;  
Matches: 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGCVOPG 8  
DB 438 RGCLRPG 444

Search completed: June 13, 2001, 14:10:42  
Job time: 143 sec



*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:43 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-9  
Perfect score: 48  
Sequence: 1 GRCCVQPC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	83.3	411	1	EF1G_TRYCR
2	38	79.2	267	1	TRV7_ANGCA
3	38	79.2	275	1	TRV2_ANGCA
4	37	77.1	277	1	TRV2_ANGCA
5	37	77.1	481	1	LBP_MOUSE
6	37	77.1	728	1	SKI_HUMAN
7	35	72.9	227	1	TRYP_SACER
8	35	72.9	525	1	UGTG_CAEEL
9	35	72.9	881	1	Y310_HUMAN
10	34	70.8	231	1	TRV2_SALSA
11	34	70.8	242	1	TRV1_SALSA
12	34	70.8	250	1	TRYP_PLEPL
13	34	70.8	357	1	KITH_HSVB6
14	34	70.8	480	1	VP5_BRD
15	34	70.8	787	1	ITB3_MOUSE
16	33	68.8	144	1	WAP3_PTG
17	33	68.8	190	1	DSR6_HUMAN
18	33	68.8	237	1	TRYP_ASTFL
19	33	68.8	256	1	HYPB_HYPLI
20	33	68.8	259	1	TRYP_STRGR
21	33	68.8	273	1	MCT7_MOUSE
22	33	68.8	273	1	MCT7_MOUSE
23	33	68.8	275	1	TRV1_MOUSE
24	33	68.8	275	1	TRV1_MOUSE
25	33	68.8	276	1	MCT6_MOUSE
26	33	68.8	404	1	NCAP_VHSVO
27	33	68.8	404	1	NCAP_VHSVO
28	33	68.8	496	1	GLK2_THEMA
29	33	68.8	591	1	ALU1_HUMAN
30	33	68.8	621	1	YC92_CAEEL
31	33	68.8	764	1	TSR_RAT
32	33	68.8	788	1	ITB3_HUMAN
33	32	66.7	116	1	SY06_MOUSE

34	32	66.7	130	1	KRA3_SHEEP	P02443	Ovis aries
35	32	66.7	131	1	KRA3_SHEEP	P02441	Ovis aries
36	32	66.7	132	1	KRA3_SHEEP	P02442	Capra hircu
37	32	66.7	278	1	EUTY_ECOLI	P77277	Escherichia
38	32	66.7	279	1	EUTY_ECOLI	P41794	Salmonella
39	32	66.7	302	1	CCDC_CAEEL	P17657	Caenorhabdi
40	32	66.7	335	1	G3P_CHUPN	O92750	Chlamydia p
41	32	66.7	443	1	MCRB_METVA	P07956	Methanococc
42	32	66.7	443	1	MCRB_METVA	P11561	Methanococc
43	32	66.7	447	1	GNT1_RABIT	P27115	Oryctolagus
44	32	66.7	505	1	GLPK_PSEAE	O51390	Pseudomonas
45	32	66.7	535	1	TLE1_BRARE	O13168	Brachydanio

## ALIGNMENTS

RESULT 1  
ID EF1G\_TRYCR STANDARD: PRT: 411 AA.  
AC P34715;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE ELONGATION FACTOR 1-GAMMA (EF-1-GAMMA).  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
NC NCBI\_TaxID:5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-Y.  
RX MEDLINE-93376525; PubMed-8367313;  
RA Billaut-Wulot O., Pommer V., Schoenack R., Pumas-Marty B.,  
RA Tabi A., Loyens M., Capron A., Ouaisel M.A.;  
RT "Nucleotide sequence of a Trypanosoma cruzi cDNA encoding a protein  
RT homologous to mammalian E1 gamma.";  
RL Nucleic Acids Res. 21:3901-3901(1993).  
CC -!- FUNCTION: PROBABLY PLAYS A ROLE IN ANCHORING THE COMPLEX TO OTHER  
CC -!- SUBUNIT: EF-1 IS COMPOSED OF FOUR SUBUNITS: ALPHA, BETA,  
CC DELTA, AND GAMMA.  
CC -!- SIMILARITY: TO OTHER SPECIES' GAMMA SUBUNIT AS WELL AS TO THE  
CC N-TERMINAL OF SOME EUKARYOTIC AMINO ACID-TRNA SYNTHETASES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: L17307; AAA02936.1; -  
DR PIR: S41648; S41648.  
DR InterPro: IPR000521; -  
DR InterPro: IPR001662; -  
DR Pfam: PF00647; EF1G-domain; 1.  
DR Pfam: PF00043; GST; 1.  
DR PROSITE: PS50040; EF1G; 1.  
KW Elongation factor; Protein biosynthesis.  
SQ SEQUENCE 411 AA; 46567 MW; F61463694B73CDB0 CRC64;

Query Match 83.3%; Score 40; DB 1; Length 411;  
Best Local Similarity 75.0%; Pred. No. 3.4;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCCVQPC 8  
DB 222 GRCCVQPC 229

RESULT 2

TRY7\_ANOGA STANDARD: PRT: 267 AA.

ID TRY7\_ANOGA AC P35041: 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE TRYPsin 7 PRECURSOR (EC 3.4.21.4).

GN TRYP7.

OS Anopheles gambiae (African malaria mosquito).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

OC Culicidae; Anopheles.

OX NCBI\_TaxID=7165;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SUAKOKO; TISSUE=Midgut;

RX MEDLINE=93327779; PubMed=8335004;

RA Mueller H.M., Crampton J.M., della Torre A., Sindén R., Crisanti A.;

RT "Members of a trypsin gene family in Anopheles gambiae are induced in the gut by blood meal."

RL EMBL J. 12:2891-2900(1993).

CC -1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD MEAL.

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- TISSUE SPECIFICITY: MIDGUT.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL; Z22930; CAAB0515.1; -

DR PIR; S40005; S40005.

DR HSSP; P00763; IDPO.

DR MEROPS; S01.112; -

DR InterPro; IPR001254; -

DR InterPro; IPR001314; -

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PROSITE; PS00134; TRYPsin\_HIS; 1.

DR PROSITE; PS00135; TRYPsin\_SER; 1.

KW Hydrolyase; Serine protease; Signal; Zymogen; Digestion;

KW Multigene family.

FT SIGNAL 1 18

FT PROPEP 19 41

FT CHAIN 42 267

FT ACT\_SITE 82 82

FT ACT\_SITE 127 127

FT ACT\_SITE 222 222

FT DISULFID 67 83

FT DISULFID 192 207

FT DISULFID 218 242

FT SITE 216 216

FT SEQUENCE 267 AA; 28456 MW; 11015871257DB678 CMC64;

Query Match

Best Local Similarity 79.2%; Score 38; DB 1; Length 267;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCVQPG 8

DB 239 GAGCAQPG 246

RESULT 3

TRY4\_ANOGA STANDARD: PRT: 275 AA.

ID TRY4\_ANOGA AC P35038: 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE TRYPsin 4 PRECURSOR (EC 3.4.21.4).

GN TRYP4.

OS Anopheles gambiae (African malaria mosquito).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

OC Culicidae; Anopheles.

OX NCBI\_TaxID=7165;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SUAKOKO; TISSUE=Midgut;

RX MEDLINE=93327779; PubMed=8335004;

RA Mueller H.M., Crampton J.M., della Torre A., Sindén R., Crisanti A.;

RT "Members of a trypsin gene family in Anopheles gambiae are induced in the gut by blood meal."

RL EMBL J. 12:2891-2900(1993).

CC -1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD MEAL.

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- TISSUE SPECIFICITY: MIDGUT.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL; Z22930; CAAB0515.1; -

DR PIR; S40005; S40005.

DR HSSP; P00763; IDPO.

DR MEROPS; S01.112; -

DR InterPro; IPR001254; -

DR InterPro; IPR001314; -

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PROSITE; PS00134; TRYPsin\_HIS; 1.

DR PROSITE; PS00135; TRYPsin\_SER; 1.

KW Hydrolyase; Serine protease; Signal; Zymogen; Digestion;

KW Multigene family.

FT SIGNAL 1 18

FT PROPEP 19 48

FT CHAIN 49 275

FT ACT\_SITE 89 89

FT ACT\_SITE 134 134

FT ACT\_SITE 230 230

FT DISULFID 74 90

FT DISULFID 199 215

FT DISULFID 226 250

FT SITE 224 224

FT SEQUENCE 275 AA; 29695 MW; 1EF29A79CE1D86C CMC64;

Query Match

Best Local Similarity 79.2%; Score 38; DB 1; Length 275;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCVQPG 8

DB 247 GAGCAQPG 254

RESULT 4

TRY2\_ANOGA STANDARD: PRT: 277 AA.



AC P35036;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TRYPsin 2 PRECURSOR (EC 3.4.21.4).  
 GN TRYP2.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Culicidae; Anopheles.  
 NC NCB1\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SUAKOKO; TISSUE=Midgeut;  
 RX MEDLINE=93327779; PubMed=8335004;  
 RA Mueller H.M., Crampston J.M., della Torre A., Crisanti A.;  
 RT "Members of a trypsin gene family in Anopheles gambiae are induced in  
 the gut by blood meal."  
 RL EMBL J. 12:2891-2900(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SUAKOKO;  
 RX MEDLINE=96101582; PubMed=7498434;  
 RA Mueller H.M., Cateruccia F., Vizioli J., della Torre A.,  
 Crisanti A.;  
 RT "Constitutive and blood meal-induced trypsin genes in Anopheles  
 gambiae."  
 RL Exp. Parasitol. 81:371-385(1995).  
 CC -1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD  
 MEAL.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: MIDGUT.  
 CC -1- DEVELOPMENTAL STAGE: INDUCED IN THE MIDGUT OF FEMALE AFTER BLOOD  
 MEAL.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sdb.ch/announce/  
 or send an email to license@isb-sdb.ch).  
 CC -----  
 DR EMBL: 218890; CAA79328.1; -;  
 DR EMBL: 222930; CAA80518.1; -;  
 DR PIR: S35413; S35413.  
 DR PIR: S35340; S35340.  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.112; -;  
 DR InterPro: IPR001254; -;  
 DR InterPro: IPR001314; -;  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Hydrolase; Serine protease; Signal; Zymogen; Digestion;  
 KW Multigene family.  
 FT SIGNAL 1 19  
 FT PROPEP 20 50  
 FT CHAIN 51 277  
 FT ACT\_SITE 91 91  
 FT ACT\_SITE 136 136  
 FT ACT\_SITE 232 232  
 FT DISULFID 76 92  
 FT DISULFID 201 217  
 FT DISULFID 228 252  
 FT SITE 226 226  
 FT SEQUENCE 277 AA; 29763 MW; 89D1C56AB4FDBD1 CRC64;

Query Match 77.1%; Score 37; DB 1; Length 277;  
 Best Local Similarity 75.0%; Pred. No. 8.4;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GRGCVPG 8  
 DB 249 GYGCAQPG 256  
 RESULT 5  
 ID LBP\_MOUSE STANDARD; PRT; 481 AA.  
 AC 061805;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE LIPOPOLYSACCHARIDE-BINDING PROTEIN PRECURSOR (LBP).  
 GN LBP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BAIB/C;  
 RX MEDLINE=97289150; PubMed=9144073;  
 RA Lengacher S., Jongeneel C.V., Le Roy D., Lee J.D., Kravchenko V.,  
 Ulevitch R.J., Glauser M.P., Heumann D.;  
 RT "Reactivity of murine and human recombinant LPS-binding protein (LBP)  
 within LPS and gram negative bacteria."  
 RL J. Inflamm. 47:165-172(1995).  
 CC -1- FUNCTION: BINDS TO THE LIPID A MOIETY OF BACTERIAL  
 LIPOPOLYSACCHARIDES (LPS), A GLYCOLIPID PRESENT IN THE OUTER  
 MEMBRANE OF ALL GRAM-NEGATIVE BACTERIA. THE LBP/LPS COMPLEX SEEMS  
 TO INTERACT WITH THE CD14 RECEPTOR.  
 CC -1- SIMILARITY: BELONGS TO THE BPI/CEP/LBP/PLRP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sdb.ch/announce/  
 or send an email to license@isb-sdb.ch).  
 CC -----  
 DR EMBL: X99347; CAA67727.1; -;  
 DR HSSP: P17213; LBP1.  
 DR MGD: MGI:1098776; Lbp.  
 DR InterPro: IPR001124; -;  
 DR Pfam: PF01273; LBP\_BPI\_CETP; 1.  
 DR PROSITE: PS00400; LBP\_BPI\_CETP; 1.  
 KW Lipid transport; Antibiotic; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 481  
 FT CAROXYD 300 300  
 FT CAROXYD 355 355  
 FT SEQUENCE 481 AA; 53312 MW; 34E9C06C9AB678 CRC64;  
 Query Match 77.1%; Score 37; DB 1; Length 481;  
 Best Local Similarity 75.0%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GRGCVPG 8  
 DB 22 GTGCVNPG 29  
 RESULT 6  
 ID SKI\_HUMAN STANDARD; PRT; 728 AA.  
 AC P12755;  
 DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, last sequence update)  
 DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DE SKI ONCOGENE (C-SKI).  
 GN SKI.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9345144; PubMed=2762147;  
 RA Nomura N., Sasamoto S., Ishii S., Matsui M., Ishizaki R.;  
 RT "Isolation of human cDNA clones of ski and the ski-related gene,  
 RT sno.";  
 RL Nucleic Acids Res. 17:5489-5500(1989).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN TERMINAL DIFFERENTIATION OF SKELETAL  
 CC MUSCLE CELLS BUT NOT IN THE DETERMINATION OF CELLS TO THE MYOGENIC  
 CC LINEAGE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SNO  
 CC ONCOGENES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X15218; CAA33288.1; -  
 DR PIR: S06053; TVHUSK.  
 DR MIM: 164780;  
 DR KW Proto-oncogene; Coiled coil; Nuclear protein; Repeat.  
 FT DOMAIN 536 710 COILED COIL (POTENTIAL);  
 SQ SEQUENCE 728 AA; 80004 MW; 9B78C4840A28C2DA CRC64;

Query Match 77.1%; Score 37; DB 1; Length 728;  
 Best Local Similarity 85.7%; Pred. No. 19;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVQP 7  
 Db 7 GRGCVQP 13

RESULT 7  
 ID TRYP\_SACER STANDARD; PRT; 227 AA.  
 AC P24664;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE TRYP-SIN (EC 3.4.21.4) (SET).  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
 CC Actinomycetales; Pseudonocardiales; Pseudonocardiales;  
 CC Saccharopolyspora.  
 CC NCBI\_TaxID=1836;  
 RN [1]  
 RP SEQUENCE.  
 RA Miyamoto K., Matsuo H., Narita K.;  
 RL (in) Dai 30 kai tanpakushitsu kouzou touronkai kouen youshishuu,  
 RL pp.77-80 (1979).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=92176171; PubMed=1794983;  
 RA Nagamine Y., Norioka S., Sakiyama F.;  
 RL unpublished results, cited by:  
 RL Yamane T., Kobuke M., Tsutsui H., Toida T., Suzuki A., Ashida T.,  
 RL Kawata Y., Sakiyama F.;  
 RL J. Biochem. 110:945-950(1991).

RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 RX MEDLINE=92176171; PubMed=1794983;  
 RA Yamane T., Kobuke M., Tsutsui H., Toida T., Suzuki A., Ashida T.,  
 RA Kawata Y., Sakiyama F.;  
 RT "Crystal structure of Streptomyces erythraeus trypsin at 2.7-A  
 RT resolution";  
 RL J. Biochem. 110:945-950(1991).  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- MISCELLANEOUS: HANDLY AUTOLYZES ITSELF AT ALL AT ITS ACTIVE PH  
 CC RANGE.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC HSSP: P00763; IDPO.  
 DR MEROPS: S01.102;-  
 DR InterPro: IPR001254;-  
 DR InterPro: IPR001314;-  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease.  
 FT ACT\_SITE 42 42 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 88 88 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 179 179 CHARGE RELAY SYSTEM.  
 FT DISULFID 27 43  
 FT DISULFID 150 164  
 FT DISULFID 175 199  
 FT SITE 173 173  
 SQ SEQUENCE 227 AA; 23308 MW; D5AC5E47B227B418 CRC64; REQUIRED FOR SPECIFICITY (BY SIMILARITY).

Query Match 72.9%; Score 35; DB 1; Length 227;  
 Best Local Similarity 62.5%; Pred. No. 16;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVQP 8  
 Db 196 GRGCVQP 203

RESULT 8  
 ID UGTG\_CAEEL STANDARD; PRT; 525 AA.  
 AC Q22295;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PUTATIVE UDP-GLUCURONOSYLTRANSFERASE UGT16 PRECURSOR (EC 2.4.1.17)  
 DE (UDP-GT).  
 GN UGT16 OR T07C5.1.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 CC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2.  
 RA McMurtry A.;  
 RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR -> UDP + ACCEPTOR  
 CC BETA-D-GLUCURONOSIDE.  
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z50006; CAA90296.1; -



TRY1\_SALSA STANDARD; PRT; 242 AA.

AC P35031;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRYPsin I PRECURSOR (EC 3.4.21.4).  
 OS Salmo salar (Atlantic salmon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
 OX NCBI\_TaxID=8030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=96035908; PubMed=7556223;  
 RA Male R., Lorens J.B., Smals A.O., Torrisen K.R.;  
 RT "Molecular cloning and characterization of anionic and cationic  
 RL variants of trypsin from Atlantic salmon.";  
 RL Eur. J. Biochem. 232:677-685(1995).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).  
 RA Smals A.O., Hordvik A.;  
 RT "Structure determination and refinement of benzamidine-inhibited  
 RL trypsin from the North Atlantic salmon (Salmo salar) at 1.82-A  
 RL resolution.";  
 RL Acta Crystallogr. D 49:318-330(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).  
 RA MEDLINE=95148588; PubMed=7846025;  
 RX Smals A.O., Helmsd E.S., Hordvik A., Willassen N.P., Male R.;  
 RT "Cold adaptation of enzymes: structural comparison between salmon and  
 RL bovine trypsins.";  
 RL Proteins 20:149-166(1994).  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
 CC TRYPsin FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X70075; CAA49680.1; -;  
 DR EMBL; X70071; CAA49676.1; -;  
 DR EMBL; X70072; CAA49677.1; -;  
 DR PIR; S31775; S31775.  
 DR PIR; S31776; S31776.  
 DR PIR; S31777; S31777.  
 DR PDB; 2TBS; 30-APR-94.  
 DR PDB; 1BTT; 01-NOV-94.  
 DR MEROPS; S01.151; -;  
 DR InterPro: IPR001254; -;  
 DR InterPro: IPR001314; -;  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PROSITE; PS00134; TRYPsin\_HIS.1.  
 DR PROSITE; PS00135; TRYPsin\_SER.1.  
 KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KM Multigene family; 3D-structure.  
 FT SIGNAL 1 15  
 FT PROPEP 16 20  
 FT CHAIN 21 242  
 FT ACT\_SITE 60 60  
 FT ACT\_SITE 104 104  
 FT ACT\_SITE 196 196  
 FT DISULFID 27 156  
 FT DISULFID 45 61  
 FT DISULFID 129 229

FT DISULFID 136 202  
 FT DISULFID 167 181  
 FT DISULFID 192 216  
 FT SITE 190 190  
 FT VARIANT 33 33  
 SQ SEQUENCE 242 AA; 25958 MW; 43F5642498067E5A CRC64;

Query Match  
 Best Local Similarity 70.8%; Score 34; DB 1; Length 242;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCVQPG 8  
 DB 213 GYGCAEPG 220

RESULT 12  
 TRYP\_PLEPL STANDARD; PRT; 250 AA.

AC P35034;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE TRYPsin PRECURSOR (EC 3.4.21.4).  
 OS Pleuronectes platessa (Plaice).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorphi; Pleuronectiformes;  
 OC Pleuronectidae; Pleuronectes.  
 OX NCBI\_TaxID=8262;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Leaver M.J., George S.G.;  
 RT Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
 CC TRYPsin FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X56744; CAA40068.1; -;  
 DR PIR; S31384; S31384.  
 DR HSSP; P00761; LEPT.  
 DR MEROPS; S01.151; -;  
 DR InterPro: IPR001254; -;  
 DR Pfam; PF00089; trypsin.1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; FALSE-NEG.  
 DR PROSITE; PS00135; TRYPsin\_SER.1.  
 KW Hydrolyase; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1 15  
 FT PROPEP 16 22  
 FT CHAIN 23 250  
 FT ACT\_SITE 62 62  
 FT ACT\_SITE 106 106  
 FT ACT\_SITE 203 203  
 FT DISULFID 29 163  
 FT DISULFID 47 63  
 FT DISULFID 133 236  
 FT DISULFID 140 209  
 FT DISULFID 174 188  
 FT DISULFID 199 223  
 FT SITE 197 197  
 SQ SEQUENCE 250 AA; 27527 MW; 637DE96185C1ABAA CRC64;

Query Match 70.8%; Score 34; DB 1; Length 250;  
Best Local Similarity 71.4%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCVOPG 7  
DB 220 GRCVOPG 226

## RESULT 13

KITH\_HSVB6 STANDARD; PRT: 357 AA.  
AC P24096;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE THYMIDINE KINASE (EC 2.7.1.21).  
CN TK.  
OS Bovine herpesvirus type 1 (strain 6660).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=103322;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89279273; PubMed=2543766;  
RA Mital S.R., Field H.J.;  
RT "Analysis of the bovine herpesvirus type 1 thymidine kinase (TK) gene from wild-type virus and TK-deficient mutants.";  
RL J. Gen. Virol. 70:901-918(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE -> ADP + THYMIDINE 5'-PHOSPHATE.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES THYMIDINE KINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D00438; BAA00339.1; -;  
DR PIR: A31330; KIBBR.  
DR HSSP: P03176; 1KTM.  
DR InterPro: IPR00189; -;  
DR Pfam: PF00693; TK\_herpes; 1.  
KM Transferase: Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 17  
FT ATP (PROBABLE).  
SQ SEQUENCE 357 AA: 36759 MW: B2099A05D5BDCD2D CRC64;

Query Match 70.8%; Score 34; DB 1; Length 357;  
Best Local Similarity 75.0%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCVOPG 8  
DB 92 GRCVOPG 99

## RESULT 14

VP5\_BRD STANDARD; PRT: 480 AA.  
AC P21230;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE OUTER CAPSID PROTEIN VP5.  
CN S5.  
OS Broadhaven virus (BRD).  
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.  
OX NCBI\_TaxID=10893;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91021056; PubMed=2171220;  
RA Moss S.R., Fukusho A., Nuttall P.A.;  
RT "RNA segment 5 of broadhaven virus, a tick-borne orbivirus, shows sequence homology with segment 5 of bluetongue virus.";  
RL Virology 179:482-484(1990).  
CC -1- FUNCTION: THE VP5 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP2) WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID.  
CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP5 FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M58030; AAA42802.1; -;  
DR PIR: A45339; A45339.  
DR InterPro: IPR000145; -;  
DR Pfam: PF00901; Orbl\_VP5; 1.  
KM Coat protein.  
SQ SEQUENCE 480 AA: 52522 MW: 28DFD78B2BE923A1 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 480;  
Best Local Similarity 71.4%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGCVOPG 8  
DB 450 RGCVOPG 456

## RESULT 15

ITB3\_MOUSE STANDARD; PRT: 787 AA.  
AC O54890;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PLATELET MEMBRANE GLYCOPROTEIN ITIA PRECURSOR (GPIIa) (INTEGRIN BETA-3) (CD61).  
GN ITGB3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/HEJ;  
RA McHugh K.P., Teitelbaum S.L., Kitazawa S., Ross F.P.;  
RT Submitted (SEP-1997) to the EMBL/Genbank/DDJ databases.  
CC -1- FUNCTION: PLATELET MEMBRANE GLYCOPROTEIN ITB-ITIA IS THE PLATELET RECEPTOR FOR ADHESIVE PROTEINS. IT BINDS FIBRINOGEN, FIBRONECTIN, VITRONECTIN AND VWF. INTEGRINS RECOGNIZE THE SEQUENCE R-G-D IN THEIR LIGAND.  
CC -1- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-3 IS EITHER ASSOCIATED WITH ALPHA-ITB (GPIIB) OR WITH ALPHA-V (VITRONECTIN RECEPTOR).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE BONDS.  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC or send an email to license@slsb.slb.ch).

DR EMBL; AF026509; AAB94086.1; -

DR MGD; MGI:96612; Itgb3.

DR InterPro: IPR000561; -

DR InterPro: IPR001169; -

DR InterPro: IPR002369; -

DR Pfam; PF00362; Integrin\_B. 1.

DR PRINTS; PR01186; INTEGRIN\_BETA. 3.

DR PROSITE; PS00243; INTEGRIN\_BETA. 3.

DR PROSITE; PS00022; EGF\_1; UNKNOWN. 2.

DR PROSITE; PS01186; EGF\_2; UNKNOWN. 1.

KM Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat;  
Extracellular matrix; Cytoskeleton; Signal.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 787 PLATELET MEMBRANE GLYCOPROTEIN IIIA.

FT DOMAIN 26 717 EXTRACELLULAR.

FT TRANSMEM 718 740 POTENTIAL.

FT DOMAIN 741 787 CYTOPLASMIC.

FT DOMAIN 462 628 CYSTEINE-RICH REPEATS.

FT REPEAT 462 510 I.

FT REPEAT 511 552 II.

FT REPEAT 553 591 III.

FT REPEAT 592 628 IV.

FT DISULFID 30 460 BY SIMILARITY.

FT DISULFID 38 48 BY SIMILARITY.

FT DISULFID 41 74 BY SIMILARITY.

FT DISULFID 51 63 BY SIMILARITY.

FT DISULFID 202 209 BY SIMILARITY.

FT DISULFID 257 298 BY SIMILARITY.

FT DISULFID 399 411 BY SIMILARITY.

FT DISULFID 431 680 BY SIMILARITY.

FT DISULFID 458 462 BY SIMILARITY.

FT DISULFID 473 485 BY SIMILARITY.

FT DISULFID 482 520 BY SIMILARITY.

FT DISULFID 487 496 BY SIMILARITY.

FT DISULFID 498 511 BY SIMILARITY.

FT DISULFID 526 531 BY SIMILARITY.

FT DISULFID 528 561 BY SIMILARITY.

FT DISULFID 533 546 BY SIMILARITY.

FT DISULFID 548 553 BY SIMILARITY.

FT DISULFID 567 572 BY SIMILARITY.

FT DISULFID 569 600 BY SIMILARITY.

FT DISULFID 574 583 BY SIMILARITY.

FT DISULFID 585 592 BY SIMILARITY.

FT DISULFID 606 611 BY SIMILARITY.

FT DISULFID 608 656 BY SIMILARITY.

FT DISULFID 613 623 BY SIMILARITY.

FT DISULFID 626 629 BY SIMILARITY.

FT DISULFID 633 642 BY SIMILARITY.

FT DISULFID 639 712 BY SIMILARITY.

FT DISULFID 660 688 BY SIMILARITY.

FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 787 AA; 86694 MW; B1570599A8C438A3 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 787;  
Best Local Similarity 50.0%; Pred. No. 70;  
Matches 8; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRG-----CVQPG 8  
||| |||||  
Db 602 GRGNCCEGSCVCVQPG 617

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:28 ; Search time 225.85 Seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-9  
Perfect score: 48  
Sequence: 1 GRGCVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP unclassified:\*  
13: SP vertebrate:\*  
14: SP virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the entry being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	85.4	2240	2 09S129	09S129 streptomyc
2	38	79.2	783	2 068952	068952 heparitis c
3	37	77.1	534	2 006418	006418 mycobacteri
4	36	75.0	265	5 09XY55	09XY55 ctenocephal
5	36	75.0	458	4 099763	099763 homo sapien
6	36	75.0	722	14 P88901	P88901 killam rat
7	36	75.0	1241	4 014148	014148 homo sapien
8	35	72.9	145	1 026242	026242 methanobact
9	35	72.9	272	2 054137	054137 saccharopol
10	35	72.9	273	4 09UHFL	09UHFL homo sapien
11	35	72.9	433	2 09RY94	09RY94 deinococcus
12	35	72.9	593	14 087093	087093 simian herp
13	35	72.9	622	5 09VWHS	09VWHS drosophila
14	35	72.9	706	2 030536	030536 pseudomonas
15	34	70.8	218	2 006603	006603 mycobacteri
16	34	70.8	247	13 09W705	09W705 paratichthy
17	34	70.8	250	13 093265	093265 pleuronecte
18	34	70.8	254	5 09XY48	09XY48 ctenocephal
19	34	70.8	261	5 09V445	09V445 drosophila

20	34	70.8	263	5 09V5X7	09V5X7 drosophila
21	34	70.8	288	2 052588	052588 mycobacteri
22	34	70.8	303	5 076900	076900 drosophila
23	34	70.8	303	5 09W586	09W586 drosophila
24	34	70.8	317	2 050655	050655 mycobacteri
25	34	70.8	331	11 P70390	P70390 mus musculi
26	34	70.8	468	10 09SUT6	09SUT6 arabidopsis
27	34	70.8	478	1 09YFD5	09YFD5 aeropyrum p
28	34	70.8	680	11 09QW15	09QW15 mus sp. bet
29	34	70.8	722	13 09PS25	09PS25 lampetra ja
30	34	70.8	723	11 09QW16	09QW16 rattus sp.
31	34	70.8	757	3 09VWFS	09VWFS drosophila
32	34	70.8	769	5 09WEN9	09WEN9 drosophila
33	34	70.8	842	10 049527	049527 arabidopsis
34	34	70.8	1088	11 09R189	09R189 rattus norv
35	34	70.8	1159	5 060981	060981 leishmania
36	34	70.8	1218	5 09VBR6	09VBR6 drosophila
37	33	68.8	57	11 064220	064220 rattus norv
38	33	68.8	81	10 09M554	09M554 ipomoea bat
39	33	68.8	81	10 09M5R6	09M5R6 ipomoea bat
40	33	68.8	159	4 09Q017	09Q017 homo sapien
41	33	68.8	162	1 09YDA2	09YDA2 aeropyrum p
42	33	68.8	197	5 09NB89	09NB89 agrotis ips
43	33	68.8	226	2 030743	030743 rhodobacter
44	33	68.8	226	2 P97177	P97177 rhodobacter
45	33	68.8	226	5 045327	045327 caenorhabdi

## ALIGNMENTS

RESULT 1  
ID 09S129 PRELIMINARY; PRT; 2240 AA.  
AC 09S129;  
DT 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DT 01-OCT-2000 (TREMUREL. 15, Last annotation update)  
DE PUTATIVE MULTI-DOMAIN BETA KETO-ACYL SYNTHASE.  
GN SCJ21.08.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_Taxid=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Parthill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Parthill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
Kinsashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL: AL109747; CAB52354.1; -;  
DR INTERPRO: IPR000794; -;  
DR INTERPRO: IPR01143; -;  
DR PFAM: PF00109; ketoacyl-synt. 2;  
DR PFM: PF01377; thioester-dehyd. 2;  
SQ SEQUENCE 2240 AA: 235991 MW: E3281D288AD44CB0 CRC64;

Query Match 85.4%; Score 41; DB 2; Length 2240;  
Best Local Similarity 87.5%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 DB 10 GRGCVLPG 17

## RESULT 2

068952 PRELIMINARY: PRT: 783 AA.  
 AC 068952: HEPATITIS C VIRUS, JK4;  
 DT 01-NOV-1996 (TREMBLER, 01, Created)  
 DT 01-NOV-1996 (TREMBLER, 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLER, 13, Last annotation update)  
 DE (JK4) CORE, E1, NS1/E2 AND NS2 GENES (FRAGMENTS).  
 GN JK4.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PATIENT 4: HEPATITIS C VIRUS, JK4;  
 RX MEDLINE=93119270; PubMed=8380322;  
 RA Honda M., Kaneo S., Ueno M., Kobayashi K., Murakami S.;  
 RT "Sequence analysis of putative structural regions of hepatitis C virus  
 RT isolated from 5 Japanese patients with hepatocellular carcinoma."  
 RL Arch. Virol. 128:163-169(1993).  
 DR EMBL: X61594; CAA43791.1; -.  
 FT CHAIN 1  
 FT NON\_CONS 191  
 FT CHAIN 192  
 FT CHAIN 192  
 FT NON\_CONS 383  
 FT CHAIN 384  
 FT CHAIN 384  
 FT NON\_CONS 733  
 FT CHAIN 734  
 FT CHAIN 734  
 FT NON\_TER 783  
 SO SEQUENCE 783 AA; 85808 MW; A3145A3D310F9ESC CRC64;

Query Match 79.2%; Score 38; DB 14; Length 783;  
 Best Local Similarity 75.0%; Pred. No. 36;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 DB 73 GRGCVOPG 80

## RESULT 3

006418 PRELIMINARY: PRT: 534 AA.  
 AC 006418: TREMBLER, 04, Created  
 DT 01-JUL-1997 (TREMBLER, 04, Last sequence update)  
 DT 01-JUL-1997 (TREMBLER, 04, Last sequence update)  
 DE HYPOTHETICAL 57.3 KDA PROTEIN.  
 GN RV0552 OR MTCY25D10.31.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z95558; CAB08963.1; -.  
 DR TUBERCULIST; RV0552; -.  
 KW Hypothetical protein.  
 SO SEQUENCE 534 AA; 57339 MW; 1AD03B04A977682 CRC64;

Query Match 77.1%; Score 37; DB 2; Length 534;  
 Best Local Similarity 75.0%; Pred. No. 39;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 DB 55 GRGCVMPG 62

## RESULT 4

09XY55 PRELIMINARY: PRT: 265 AA.  
 AC 09XY55: TREMBLER, 12, Created  
 DT 01-NOV-1999 (TREMBLER, 12, Last sequence update)  
 DT 01-NOV-1999 (TREMBLER, 12, Last sequence update)  
 DE TRYPSIN-LIKE SERINE PROTEASE.  
 GN SP-28.  
 OS Ctenocephalides felis (Cat flea).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Siphonaptera; Pulicidae;  
 CC Ctenocephalides.  
 OX NCBI\_TaxID=7515;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99124360; PubMed=9927170;  
 RA Gaines P.J., Sampson C.M., Rushlow K.E., Stiegler G.L.;  
 RT "Cloning of a family of serine protease genes from the cat flea  
 RT Ctenocephalides felis."  
 RL Insect Mol. Biol. 81:11-22(1999).  
 DR EMBL: AF053913; AAD21833.1; -.  
 DR HSSP: P00763; IDPO.  
 DR INTERPRO: IPR001254; -.  
 DR INTERPRO: IPR001314; -.  
 DR PFM: PF00089; trypsin\_1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Protease.  
 SO SEQUENCE 265 AA; 28486 MW; 17DFA43A82A5086D CRC64;

Query Match 75.0%; Score 36; DB 5; Length 265;  
 Best Local Similarity 62.5%; Pred. No. 31;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVOPG 8  
 DB 235 GGGCVRPG 242

## RESULT 5

099763 PRELIMINARY: PRT: 458 AA.  
 AC 099763: TREMBLER, 03, Created  
 DT 01-MAY-1997 (TREMBLER, 03, Last sequence update)  
 DT 01-MAY-1997 (TREMBLER, 03, Last sequence update)  
 DE SERINE/THREONINE PROTEIN KINASE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]



```
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA MEDLINE-96207227; PubMed-8619474;
RT Anderson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT "A 'double adaptor' method for improved shotgun library
  construction.";
RL Anal. Biochem. 236:107-113(1996).
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA Yu W., Anderson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
  Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U79240; AAB50198.1; -.
DR INTERPRO: IPR000719; -.
DR INTERPRO: IPR002290; -.
DR PFWAM: PF00069; PKinase; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Serine/threonine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 458 AA; 51167 MW; D13C34888236AE20 CRC64;

Query Match 75.0%; Score 36; DB 4; Length 458;
Best Local Similarity 62.5%; Pred. No. 51;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVQPG 8
DB 444 GCGCLHPG 451

RESULT 6
P88901 PRELIMINARY; PRT; 722 AA.
AC P88901;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE NON-CAVSD PROTEIN (NS1), (NS2), (VPI), AND (VP2) GENES, COMPLETE CDS.
GN VPI.
OS Kilham rat virus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=12441;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCI:
RA Brown D.W., Like A.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U79033; AAB38327.1; -.
DR HSSP: P07302; IMVM.
DR INTERPRO: IPR001403; -.
DR PFWAM: PF00740; Parvcoat; 1.
DE PFAM: PF00740; Parvcoat; 1.
SQ SEQUENCE 722 AA; 79670 MW; 95A659BA3090C51D CRC64;

Query Match 75.0%; Score 36; DB 14; Length 722;
Best Local Similarity 85.7%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGCVQPG 8
DB 10 RGCVQPG 16

RESULT 7
Q1418 PRELIMINARY; PRT; 1241 AA.
ID Q1418
AC Q1418;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
```

```
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE KIA0135 PROTEIN (FRAGMENT).
GN KIA0135.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-96127530; PubMed-8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
  The coding sequences of 40 new genes (K1A0121-K1A0160) deduced by
  analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
RN [2]
RP SEQUENCE OF 1-54 FROM N.A.
RA Ceulemans H., Van Eynde A., Perez-Callejon E., Stalmans W., Bollen M.;
RT "Structure and splice products of the human gene encoding sds22, a
  putative mitotic regulator of protein phosphatase-1.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: D50925; BAA09484.1; -.
DR EMBL: AF067137; AAC23506.1; -.
DR INTERPRO: IPR000014; -.
DR INTERPRO: IPR000719; -.
DR INTERPRO: IPR001245; -.
DR INTERPRO: IPR002290; -.
DR PFWAM: PF00069; PKinase; 1.
DR PFWAM: PF00089; PAS; 3.
DR PRINTS: PR00109; TYRKINASE.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1241 AA; 134103 MW; B651937986664A84 CRC64;

Query Match 75.0%; Score 36; DB 4; Length 1241;
Best Local Similarity 62.5%; Pred. No. 1,3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVQPG 8
DB 1227 GCGCLHPG 1234

RESULT 8
O26242 PRELIMINARY; PRT; 145 AA.
ID O26242
AC O26242;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HETEROIDISULFIDE REDUCTASE HDRO RELATED PROTEIN.
GN MTH139.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
OX NCBI_TaxID=2166;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE-98037514; PubMed-9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
  Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
  Harrison D., Hoang L., Keagle P., Lunn W., Pothier B., Qiu D.,
  Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
  Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
  RA McDougall S., Shimer G., Goyal A., Pletovsk S., Church G.M.,
  RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
  deltaH: functional analysis and comparative genomics.";
```

RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL: AE000803; AAB84645.1; -  
 DR INTERPRO: IPR000194; -  
 DR PROSITE: PS00152; ATPASE\_ALPHA\_BETA; UNKNOWN\_1.  
 SQ SEQUENCE 145 AA; 16448 MW; 694CD4CE1A6560A8 CRC64;

Query Match 72.9%; Score 35; DB 1; Length 145;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRCVOP 7  
 111111  
 Db 100 GGCVCOP 106

RESULT 9  
 ID 054137 PRELIMINARY; PRT; 272 AA.

AC 054137;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE TRYPsin-LIKE PROTEASE.  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;  
 CC Saccharopolyspora.  
 OX NCBI\_TaxID=1836;  
 OX 11  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96064153; PubMed=8543568;  
 RA "Nagamine-Matsuka Y., Norioke S., Sakiyama F.;  
 RT "Molecular cloning, nucleotide sequence, and expression of the gene  
 RT encoding a trypsin-like protease from Streptomyces erythraeus.";  
 RL J. Biochem. 118:338-346(1995).  
 DR EMBL: D30760; BAA06423.1; -  
 DR HSSP: P00763; IDPO.  
 DR INTERPRO: IPR001254; -  
 DR INTERPRO: IPR001314; -  
 DR PFM: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Protease.  
 SQ SEQUENCE 272 AA; 27793 MW; C309A0F0B9B909A CRC64;

Query Match 72.9%; Score 35; DB 2; Length 272;  
 Best Local Similarity 62.5%; Pred. No. 49;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRCVOP 8  
 111111  
 Db 238 GGCVCOP 245

RESULT 10  
 ID 09UHFL PRELIMINARY; PRT; 273 AA.

AC 09UHFL;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE NOTCH4-LIKE PROTEIN.  
 GN ZNFU1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 OX 11  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RA Sheppard P., Jelinek L., Whitmore T., Blumberg H., Lehner J.,

RA O'Hara P.;  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF166111; AAF0429.1; -  
 DR HSSP: P35555; IEMN.  
 DR INTERPRO: IPR000152; -  
 DR INTERPRO: IPR000561; -  
 DR INTERPRO: IPR001881; -  
 DR PFM: PF00008; EGF; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 SQ SEQUENCE 273 AA; 29617 MW; 5740BB845ED5A988 CRC64;

Query Match 72.9%; Score 35; DB 4; Length 273;  
 Best Local Similarity 75.0%; Pred. No. 49;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRCVOP 8  
 111111  
 Db 114 GGCVCOP 121

RESULT 11  
 ID 09RY94 PRELIMINARY; PRT; 433 AA.  
 AC 09RY94;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE BENZOATE MEMBRANE TRANSPORT PROTEIN, PUTATIVE.  
 GN DR0056.  
 OS Deinococcus radiodurans.  
 CC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 OX 11  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI;  
 RA MDL=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome Sequence of the Radioreistant Bacterium Deinococcus  
 RT radiodurans RI.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AE001868; AAF09646.1; -  
 DR TIGR: DR0056; -  
 DR INTERPRO: IPR000566; -  
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN\_1.  
 SQ SEQUENCE 433 AA; 44251 MW; 00D82E409EFE0396 CRC64;

Query Match 72.9%; Score 35; DB 2; Length 433;  
 Best Local Similarity 75.0%; Pred. No. 75;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRCVOP 8  
 111111  
 Db 383 GRCVOP 390

RESULT 12  
 ID 087093 PRELIMINARY; PRT; 593 AA.

AC 087093;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)

```

DE GLYCOPROTEIN G (HOMOLOGUE OF HSV-2 USA).
OS Simian herpes B virus (Cercopithecid herpesvirus 1) (Shbv).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
OC Alphaherpesvirinae; Simplexvirus.
ON NCBI_TaxID=10325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CYNO 2.
RX MEDLINE=9600307; PubMed=7561753;
RA Slomka M.J., Harrington L., Arnold C., Norcott J.P.N., Brown D.W.;
RT "Complete nucleotide sequence of the herpesvirus simiae glycoprotein G
RT gene and its expression as an immunogenic fusion protein in
RT bacteria."
RL J. Gen. Virol. 76:2161-2168(1995).
DR EMBL: Z46268; CA86431.1;
SQ SEQUENCE 593 AA; 60731 MW; 79BD266FA82B8B5 CRC64;

Query Match 72.9%; Score 35; DB 14; Length 593;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GRGCVOPG 8
Db 248 GRGCVSPG 255

RESULT 13
O9VMH5 PRELIMINARY; PRT; 622 AA.
AC O9VMH5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE KR-H1 PROTEIN.
GN KR-H1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Plamkoc C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fossler G., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Platan G.S., Pan S., Pollard J., Puft V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```

```

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003612; AAF52343.1;
DR HSSP: P08047; ISP2.
DR FLYBASE: FBgn0028420; Kr-h1.
DR INTERPRO: IPR000822; Kr-h1.
DR PFAM: PF00096; ZF-C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 6.
SQ SEQUENCE 622 AA; 68241 MW; 01E2B29ED64B9E8B CRC64;

Query Match 72.9%; Score 35; DB 5; Length 622;
Best Local Similarity 62.5%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GRGCVOPG 8
Db 554 GGGLCPG 561

RESULT 14
O30536 PRELIMINARY; PRT; 706 AA.
AC O30536;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PCRD.
GN PCRD.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
ON NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-388;
RX MEDLINE=98037517; PubMed=9371466;
RA Yahr T.L., Mende-Kueller L.M., Friese M.B., Frank D.W.;
RT "Identification of type III secreted products of the Pseudomonas
RT aeruginosa exoenzyme S regulon."
RL J. Bacteriol. 179:7165-7168(1997).
DR EMBL: AF010150; AAC45944.1;
DR INTERPRO: IPR001712;
DR PFAM: PF00771; FHIPEP; 1.
DR PRINTS: PR00949; TYPE3IMAPROT.
DR PROSITE: PS00994; FHIPEP; 1.
SQ SEQUENCE 706 AA; 77009 MW; 1BA8BD2DF2B46491 CRC64;

Query Match 72.9%; Score 35; DB 2; Length 706;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 RGCVOPG 8
Db 438 RGCLRP 444

RESULT 15
O06603 PRELIMINARY; PRT; 218 AA.
AC O06603;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)

```

DE HYPOTHETICAL 23.3 KDA PROTEIN (FRAGMENT).  
GN RV1587C OR MTCY336.17.  
OS Mycobacterium tuberculosis.  
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RM [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Sultson J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
DR EMBL; Z95586; CAB09082.1; -.  
DR TUBERCULIST; RV1587c; -.  
KW Hypothetical protein..  
FT NON\_TER 218  
SQ SEQUENCE 218 AA; 23296 MW; 66376488F3994DC6 CRC64;

Query Match 70.8%; Score 34; DB 2; Length 218;  
Best Local Similarity 71.4%; Pred. No. 62;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGCVP8 8  
|||:||  
Db 146 RGCCTPG 152

Search completed: June 13, 2001, 14:20:29  
Job time: 729 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:33 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-9

Perfect score: 48

Sequence: 1 GRGCVPG 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

## Database :

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	79.2	246	1	US-07-843-125-11
2	36	75.0	267	3	US-08-906-769-145
3	36	75.0	267	4	US-08-906-616-145
4	36	75.0	267	4	US-08-639-075A-145
5	36	75.0	267	4	US-09-004-731-67
6	36	75.0	267	4	US-09-012-431-145
7	36	75.0	267	4	US-08-749-699-67
8	34	70.8	254	4	US-08-906-769-129
9	34	70.8	254	4	US-08-906-616-129
10	34	70.8	254	4	US-08-639-075A-129
11	34	70.8	254	4	US-09-012-431-129
12	33	68.8	58	2	US-08-465-380-60
13	33	68.8	58	2	US-08-486-397-60
14	33	68.8	58	2	US-08-486-399-60
15	33	68.8	58	2	US-08-461-965-60
16	33	68.8	58	2	US-08-634-641-60
17	33	68.8	58	3	US-09-249-471-60
18	33	68.8	58	3	US-09-249-472-60
19	33	68.8	58	3	US-09-249-451-60
20	33	68.8	58	3	US-08-809-455-60
21	33	68.8	58	3	US-09-249-461-60
22	33	68.8	58	4	US-09-249-448-60
23	33	68.8	76	2	US-08-465-380-29
24	33	68.8	76	2	US-08-480-478-58
25	33	68.8	76	2	US-08-486-397-29
26	33	68.8	76	2	US-08-486-399-29
27	33	68.8	76	2	US-08-461-965-29

28	33	68.8	76	2	US-08-326-110A-58	Sequence 58, Appl
29	33	68.8	76	2	US-08-634-641-29	Sequence 29, Appl
30	33	68.8	76	3	US-09-249-471-29	Sequence 29, Appl
31	33	68.8	76	3	US-09-249-472-29	Sequence 29, Appl
32	33	68.8	76	3	US-09-249-451-29	Sequence 29, Appl
33	33	68.8	76	3	US-08-809-455-29	Sequence 29, Appl
34	33	68.8	76	3	US-09-249-461-29	Sequence 29, Appl
35	33	68.8	76	4	US-09-249-448-29	Sequence 29, Appl
36	33	68.8	92	2	US-07-728-215-39	Sequence 39, Appl
37	33	68.8	92	2	US-07-728-215-41	Sequence 41, Appl
38	33	68.8	116	3	US-08-545-809A-143	Sequence 143, Appl
39	33	68.8	223	1	US-08-278-091-13	Sequence 13, Appl
40	33	68.8	223	1	US-08-483-855-13	Sequence 13, Appl
41	33	68.8	223	1	US-08-472-173-13	Sequence 13, Appl
42	33	68.8	223	2	US-08-487-167-13	Sequence 13, Appl
43	33	68.8	223	2	US-08-482-816-13	Sequence 13, Appl
44	33	68.8	223	2	US-08-296-149-13	Sequence 13, Appl
45	33	68.8	223	2	US-08-801-499-13	Sequence 13, Appl

## ALIGNMENTS

RESULT 1  
Sequence 11, Application US/07843125  
Patent No. 5395750  
GENERAL INFORMATION:  
APPLICANT: Dillion, Patrick J  
TITLE OF INVENTION: Methods for Producing Proteins which  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: George M. Gould, Esq., Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110-1199  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/843,125  
FILING DATE: 19920228  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Catherine R  
REGISTRATION NUMBER: 34240  
REFERENCE/DOCKET NUMBER: 8594  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-6208  
TELEFAX: (201) 235-3500  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-843-125-11

Query Match 79.2%; Score 38; DB 1; Length 246;  
Best Local Similarity 87.5%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GRGCVPG 8  
Db 8 GRGLVPG 15

RESULT 2  
US-08-906-769-145  
Sequence 145, Application US/08906769  
Patent No. 6077687  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Kennedy, Gary  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS: Rose & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,769  
FILING DATE:  
CLASSIFICATION:  
PRIORITY INFORMATION: DATA: US 08/639,075  
PRIORITY NUMBER: 08/639,075  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 145:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-906-769-145  
Query Match 75.0%; Score 36; DB 3; Length 267;  
Best local similarity 62.5%; Pred. No. 79;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GRGCVOPG 8  
DB 237 GGGCARRG 244  
RESULT 3  
US-08-906-616-145  
Sequence 145, Application US/08906616  
Patent No. 6121035  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Kennedy, Gary  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF

NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,616  
FILING DATE: 05-AUG-1997  
CLASSIFICATION:  
PRIORITY INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-0223  
TELEFAX: (303) 863-9700  
INFORMATION FOR SEQ ID NO: 145:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-906-616-145  
Query Match 75.0%; Score 36; DB 4; Length 267;  
Best local similarity 62.5%; Pred. No. 79;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GRGCVOPG 8  
DB 237 GGGCARRG 244  
RESULT 4  
US-08-639-075A-145  
Sequence 145, Application US/08639075A  
Patent No. 6150125  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Kennedy, Gary  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,075A  
FILING DATE: 24-APR-1996  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 145:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-639-075A-145

Query Match 75.0%; Score 36; DB 4; Length 267;  
Best Local Similarity 62.5%; Pred. No. 79;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVOPG 8  
1:11:11  
DB 237 GOGCARG 244

RESULT 5  
US-09-004-731-67  
Sequence 67, Application US/09004731  
Patent No. 6177258  
GENERAL INFORMATION:  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,731  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,699  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-004-731-67  
Query Match 75.0%; Score 36; DB 4; Length 267;  
Best Local Similarity 62.5%; Pred. No. 79;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GRGCVOPG 8  
1:11:11  
DB 237 GOGCARG 244

RESULT 6  
US-09-012-431-145  
Sequence 145, Application US/09012431  
Patent No. 6180383  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,431  
FILING DATE: 23-Jan-1998  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 145:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 145:  
US-09-012-431-145  
Query Match 75.0%; Score 36; DB 4; Length 267;  
Best Local Similarity 62.5%; Pred. No. 79;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GRGCVOPG 8  
1:11:11  
DB 237 GOGCARG 244  
RESULT 7  
US-08-749-699-67  
Sequence 67, Application US/08749699  
Patent No. 6210920  
GENERAL INFORMATION:  
APPLICANT: Wu Hunter, Shirley

APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,699  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-749-699-67

Query Match 75.0%; Score 36; DB 4; Length 267;  
Best Local Similarity 62.5%; Pred. No. 79;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVPG 8  
DB 237 GGGCARPG 244

RESULT 8  
US-08-906-769-129  
Sequence 129; Application US/08906769  
Patent No. 6077687  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,769  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 129:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-906-769-129

Query Match 70.8%; Score 34; DB 3; Length 254;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RGCVPG 8  
DB 225 RGCARPG 231

RESULT 9  
US-08-906-616-129  
Sequence 129; Application US/08906616  
Patent No. 6121035  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,616  
FILING DATE: 05-AUG-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 129:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid



TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-906-616-129

Query Match 70.8%; Score 34; DB 4; Length 254;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RCGVOPG 8  
DB 225 RCGARPG 231

RESULT 10  
US-08-639-075A-129  
Sequence 129, Application US/08639075A  
Patent No. 6150125

GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,075A  
FILING DATE: 24-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 129:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-639-075A-129

Query Match 70.8%; Score 34; DB 4; Length 254;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RCGVOPG 8  
DB 225 RCGARPG 231

RESULT 11  
US-09-012-431-129  
Sequence 129, Application US/09012431

Patent No. 6180383

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US 08/639,075

FILING DATE: 23-Jan-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/639,075

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 129:

SEQUENCE CHARACTERISTICS:

LENGTH: 254 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 129:

US-09-012-431-129

Query Match 70.8%; Score 34; DB 4; Length 254;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RCGVOPG 8  
DB 225 RCGARPG 231

RESULT 12  
US-08-465-380-60  
Sequence 60, Application US/08465380  
Patent No. 5863894

GENERAL INFORMATION:

APPLICANT: George P. Vlasuk, Patric H. Stanssens,

APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,

APPLICANT: Yves R. Laroche, Laurent S. Jespers,

APPLICANT: Yannick G.J. Ganssemans, Matthew Moyle,

APPLICANT: Peter W. Bergum

TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
PROTEIN

NUMBER OF SEQUENCES: 356

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street

STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,380  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Heligmosomoides polygyrus  
US-08-465-380-60

Query Match 68.8%; Score 33; DB 2; Length 58;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RGCVPQPG 8  
Db 48 KGCVPAPG 54

RESULT 13  
US-08-486-397-60  
Sequence 60, Application US/08486397  
Patent No. 5866542  
GENERAL INFORMATION:  
APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Ganseman, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 357  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,397  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/269  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Heligmosomoides polygyrus  
US-08-486-397-60

Query Match 68.8%; Score 33; DB 2; Length 58;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RGCVPQPG 8  
Db 48 KGCVPAPG 54

RESULT 14  
US-08-486-399-60  
Sequence 60, Application US/08486399  
Patent No. 5866543  
GENERAL INFORMATION:  
APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Ganseman, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,399  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/270

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Heligmosomoides polygyrus  
US-08-486-399-60

Query Match 68.8%; Score 33; DB 2; Length 58;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGCVOPG 8  
:11111  
DB 48 KGCVA PG 54

RESULT 15  
US-08-461-965-60  
Sequence 60, Application US/08461965  
Patent No. 5872098  
GENERAL INFORMATION:  
APPLICANT: George P. Vlausk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,965  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 210/243  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:

ORGANISM: Heligmosomoides polygyrus  
US-08-461-965-60

Query Match 68.8%; Score 33; DB 2; Length 58;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RGCVOPG 8  
:11111  
DB 48 KGCVA PG 54

Search completed: June 13, 2001, 14:16:34  
Job time: 495 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:34 ; Search time 229.28 Seconds  
(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-10

Perfect score: 47  
Sequence: 1 GRGCVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq\_0401:\*

- 1: /SID6/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SID6/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SID6/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SID6/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SID6/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SID6/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SID6/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SID6/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SID6/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SID6/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SID6/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SID6/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SID6/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SID6/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SID6/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SID6/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SID6/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SID6/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SID6/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SID6/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SID6/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SID6/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47	100.0	8	21	Y79114	Peptide antagonist
2	40	85.1	8	21	Y79113	Peptide antagonist
3	40	85.1	8	21	Y79126	Peptide antagonist
4	38	80.9	8	21	Y79106	Peptide antagonist
5	37	78.7	8	21	Y79110	Peptide antagonist
6	37	78.7	8	21	Y79116	Peptide antagonist
7	37	78.7	399	13	R20006	Zonula occludens t
8	35	74.5	346	20	W85739	Polypeptide with t
9	35	74.5	346	21	B33441	Human PRO1007 prot
10	35	74.5	346	21	Y66679	Membrane-bound pro
11	35	74.5	346	22	B50958	Human PRO1007 prot

12	34	72.3	448	15	R58666	Rat PACAP receptor
13	34	72.3	467	15	R58656	Rat PACAP receptor
14	34	72.3	476	15	R58667	Rat PACAP receptor
15	34	72.3	495	15	R58658	Rat PACAP receptor
16	33	70.2	8	21	Y79125	Peptide antagonist
17	33	70.2	67	21	Y97121	Tau conotoxin prot
18	33	70.2	336	20	Y35244	C. pneumoniae prot
19	33	70.2	623	21	B25493	Eucalyptus grandis
20	33	70.2	678	21	B10623	S. cerevisiae tran
21	33	70.2	693	21	Y79179	Haematopoietic ste
22	32	68.1	92	19	Y20611	Human neurofilamen
23	32	68.1	149	21	B40789	Human ORFX ORF53
24	32	68.1	228	20	Y35642	Amino acid sequenc
25	32	68.1	233	21	G49283	Arabidopsis thalia
26	32	68.1	270	21	G49282	Arabidopsis thalia
27	32	68.1	272	21	G49281	Arabidopsis thalia
28	32	68.1	293	20	Y27607	Human connective t
29	32	68.1	335	21	Y59247	Human WISP-1 varia
30	32	68.1	345	20	Y17652	Human putative mat
31	32	68.1	345	20	Y17640	Human WISP-1 varia
32	32	68.1	345	20	Y17642	Human WISP-1 varia
33	32	68.1	345	20	Y17643	Human WISP-1 varia
34	32	68.1	345	20	Y17646	Mouse putative mat
35	32	68.1	367	20	Y17647	Mouse WISP-1 prote
36	32	68.1	367	20	Y17653	Human WISP-1 varia
37	32	68.1	367	20	Y17641	Human WISP-1 prote
38	32	68.1	367	20	Y17644	Human WISP-1 varia
39	32	68.1	367	20	Y17645	Human WISP-1 varia
40	32	68.1	373	21	B50975	Human WISP-1 varia
41	32	68.1	373	21	G20684	Human PRO542 prote
42	32	68.1	382	21	G20684	Arabidopsis thalia
43	32	68.1	382	21	G42707	Arabidopsis thalia
44	32	68.1	389	21	G20683	Arabidopsis thalia
45	32	68.1	398	21	G06474	Arabidopsis thalia

#### ALIGNMENTS

RESULT	1
ID	Y79114
Y79114	standard; Peptide; 8 AA.
XX	
AC	Y79114;
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	Peptide antagonist of zonulin.
XX	
KW	zonulin; antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antidiabetic; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulvar; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
XX	
OS	Synthetic.
XX	
PN	WC200007609-A1.
XX	
PD	17-FEB-2000.
XX	
PF	28-JUL-1999; 99WO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fasano A;
XX	
DR	WPI; 2000-205565/18.
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1; Page 44; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

SO Sequence 8 AA;

Query Match 100.0%; Score 47; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8  
Db 1 grgcvcpg 8

RESULT 2  
Y79113  
ID Y79113 standard; Peptide: 8 AA.

XX Y79113;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
KM blood-brain barrier; antiinflammatory; cerebroprotective;  
KM neuroprotective; dermatological; antiulcer; antiviral;  
KM antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KM gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PE 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.

PA (UYMA-) UNIV MARYLAND BALTIMORE.

XX

PI Fasano A;  
XX  
DR WPI; 2000-205565/18.

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
XX  
PS Claim 1; Page 43; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
CC (Z), one of 25 such peptides (see Y79105-29) of the invention,  
CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
CC physiologically modulate the opening of mammalian tight junctions  
CC (TJ). The peptide antagonists are based on a common motif of ZOT  
CC and human zonulins, which is believed to be critical for receptor  
CC binding. They can be prepared by chemical synthesis or by use of  
CC recombinant DNA techniques. The peptide antagonists are used as an  
CC antiinflammatory agents in the treatment of gastrointestinal  
CC inflammation, where they bind to the ZOT receptor in the intestine  
CC and yet does not physiologically modulate the opening of TJ in the  
CC intestine. Gastrointestinal inflammation conditions give rise to  
CC increased intestinal permeability and the peptide is useful for  
CC treating intestinal conditions that cause protein losing enteropathy  
CC caused by infection, e.g. Clostridium difficile infection,  
CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
CC infestation, bacterial overgrowth, whipple's disease, diseases with  
CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
CC collagenous colitis, inflammatory bowel disease, diseases marked by  
CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
CC correction of congenital heart disease with Fontan's operation,  
CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
CC e.g. systemic lupus erythematosus or food allergies, primarily to  
CC milk.

SO Sequence 8 AA;

Query Match 85.1%; Score 40; DB 21; Length 8;  
Best Local Similarity 87.5%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8  
Db 1 grgcvcpg 8

RESULT 3

Y79126  
ID Y79126 standard; Peptide: 8 AA.

XX Y79126;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

XX Zonulin; antagonist; zonula occludens toxin receptor;  
KM blood-brain barrier; antiinflammatory; cerebroprotective;  
KM neuroprotective; dermatological; antiulcer; antiviral;  
KM antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
KM hypotensive; immunosuppressive; antiparasitic; vasotropic;  
KM gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PE 28-JUL-1999; 99WO-US16683.

XX

**Synthetic.**

KM zonulin; antagonist; zonula occludens toxin receptor;  
 KM blood-brain barrier; antiinflammatory; cerebroprotective;  
 KM neuroprotective; dermatological; antiulcer; antiviral;  
 KM antibacterial; cytosstatic; anti-HIV; vulnereary; antiallergic;





AC R20006;  
 XX  
 DT 31-MAR-1992 (first entry)  
 XX  
 DE Zonula occludens toxin.  
 XX  
 KM ZOT; cholera; vaccine; enterotoxin; diarrhoea.  
 OS  
 XX  
 PN MO9118979-A.  
 PD 12-DEC-1991.  
 XX  
 PF 05-JUN-1991; 91MO-US03812.  
 XX  
 PR 05-JUN-1990; 90US-0533315.  
 XX  
 PA (UYMA-) UNIV MARYLAND BALTI.  
 XX  
 PI Kaper JB, Baudry-Maurell B, Fasano A;  
 XX  
 DR WPI: 1992-007465/01.  
 DR N-PSDB: Q20185.  
 XX  
 PT New Vibrio cholerae strains - comprise restriction endonuclease  
 fragment encoding toxin, used as vaccines against cholera  
 XX  
 PS Disclosure; Fig 18; 83pp; English.  
 XX  
 CC The amino acid sequence is that of the zonula occludens toxin (ZOT).  
 CC It may be responsible for diarrhoea in some strains of cholera and  
 CC the ZOT gene or fragments of it are deleted from strains of Vibrio  
 CC cholerae (V.C.) to be used as vaccines. These V.C. strains have 100%  
 CC efficacy in protecting humans against subsequent infection with a  
 CC strain of a similar serotype and avoid undesirable side effects such  
 CC as diarrhoea, nausea and cramping. Cultures of these strains may be  
 CC used for prodn. of vaccines against cholera.  
 CC  
 SQ Sequence 399 AA;  
 XX

Query Match 78.7%; Score 37; DB 13; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GRGCVQDG 8  
 II IIIII  
 Db 291 grlcvgdg 298

RESULT 8  
 W85739 ID W85739 standard; Protein; 346 AA.  
 XX  
 AC W85739;  
 XX  
 DT 01-OCT-1999 (first entry)  
 XX  
 DE Polypeptide with transmembrane domain.  
 XX  
 KM Transmembrane protein; transmembrane domain; translocation protein;  
 KM endoplasmic reticulum; sec61 alpha subunit; research; probe;  
 KM diagnosis; gene therapy; nutritional supplement;  
 KM cytokine proliferation; cell proliferation; cell differentiation;  
 KM haematopoiesis regulation; tissue growth; activin; inhibin;  
 KM chemotactic activity; chemokinesis; haemostasis; thrombolysis;  
 KM receptor; ligand; anti-inflammatory; tumour inhibition;  
 KM antibody production.  
 XX  
 OS Homo sapiens.  
 OS  
 PN MO9918202-A2.

XX  
 PD 15-APR-1999.  
 XX  
 PF 05-OCT-1998; 98MO-JP04474.  
 XX  
 PR 08-OCT-1997; 97JP-0276269.  
 XX  
 PA (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENT.  
 XX  
 PI Kato S, Sekine S;  
 XX  
 DR WPI: 1999-277267/23.  
 DR N-PSDB: X08740, X08746.  
 XX  
 PT Human transmembrane proteins and nucleotide sequences  
 PS Claim 1; Page 73-75; 96pp; English.  
 XX  
 CC A novel method of cloning cDNAs from the human full-length cDNA  
 CC bank is described. The protein is identified by the presence of a  
 CC hydrophobic N-terminal secretory signal region, knowledge of the  
 CC protein function is not required, as in e.g. methods of expression  
 CC cloning. The cDNAs encode proteins comprising transmembrane domains  
 CC and can be used as probes for gene diagnosis and gene sources for  
 CC gene therapy. Abnormalities of membrane proteins are associated  
 CC with a number of hitherto-cryptogenic diseases. e.g. cystic  
 CC fibrosis. Polynucleotides and proteins of the invention can be used  
 CC in research; as nutritional supplements; for cell proliferation;  
 CC cell differentiation or cytokine activity or to induce the  
 CC production of cytokines; for immune stimulating or suppressing  
 CC activity i.e. generation of antibodies; for haematopoiesis regulating  
 CC activity; for tissue growth activity; for activin/inhibin activity;  
 CC for chemotactic/chemokinetic activity; for haemostatic and  
 CC thrombolytic activity; for receptor/ligand activity; for  
 CC anti-inflammatory activity or for tumour inhibition activity.  
 CC A search of a protein database revealed no proteins having an  
 CC analogy with this sequence.  
 CC  
 SQ Sequence 346 AA;  
 XX

Query Match 74.5%; Score 35; DB 20; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 RGCYQD 7  
 IIIII  
 Db 191 rgcvqgd 196

RESULT 9  
 B33441 ID B33441 standard; Protein; 346 AA.  
 XX  
 AC B33441;  
 XX  
 DT 29-JAN-2001 (first entry)  
 XX  
 DE Human PRO1007 protein UNQ491 SEQ ID NO:146.  
 XX  
 KM Human; immune related disease; diagnosis; antiinflammatory; cardiant;  
 KM dermatological; antiarthritic; antirheumatic; immunosuppressive;  
 KM haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;  
 KM antiaanaemic; hepatotropic; virucide; antiparasitic; antiallergic;  
 KM antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;  
 KM osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;  
 KM idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;  
 KM systemic vasculitis; autoimmune haemolytic anemia; diabetes mellitus;  
 KM autoimmune thrombocytopenia; immune-mediated renal disease;  
 KM demyelinating disease; hepatobiliary disease; Whipple's disease;  
 KM inflammatory bowel disease; gluten-sensitive enteropathy;  
 KM autoimmune disease; immune-mediated skin disease; allergic disease;



```

PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 19-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 23-JUN-1998; 98US-0090354.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090576.
PR 25-JUN-1998; 98US-0090578.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 02-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.

```

```

PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096737.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-SEP-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

```

(GETH ) GENENTECH INC.

Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;  
 Wood WI, Yuan J;

WPI: 2000-072883/06.  
 N-PSDB: 265016.

Membrane-bound proteins and related nucleotide sequences -

claim 12; Fig 125; 822pp; English.

CC The invention provides membrane-bound PRO polypeptides and  
 CC polynucleotides encoding them. The PRO sequences of the invention were  
 CC identified based on extracellular domain homology screening. The PRO  
 CC sequences have homology with proteins including LDL receptors, TIR  
 CC ligands and various enzymes. The membrane-bound proteins and receptor  
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor  
 CC immunoadhesins, for instance, can be used as therapeutic agents to block  
 CC receptor-ligand interactions. The membrane-bound proteins can also be  
 CC employed for screening of potential peptide or small molecule inhibitors  
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences  
 CC are useful as hybridization probes, in chromosome and gene mapping and in  
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences  
 CC will also be useful for the preparation of PRO polypeptides, especially  
 CC by recombinant techniques.

XX Sequence 346 AA;

Query Match 74.5%; Score 35; DB 21; Length 346;

Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RGCYOD 7  
 |||||

Db 191 rgcvqd 196

## RESULT 11

B50958  
ID B50958 standard; Protein: 346 AA.

XX AC B50958;

XX DT 21-MAR-2001 (first entry)

XX DE Human PRO1007 protein.

XX KW Human; PRO: cytostatic; neuroprotective; respiratory general;  
KW antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;  
KW PRO agonist; cancer; inflammatory disorder; immunological disorder.

XX OS Homo sapiens.

XX PN WO200073348-A2.

XX PD 07-DEC-2000.

XX PF 30-MAY-2000; 2000MO-US14941.

XX PR 02-JUN-1999; 99MO-US12252.

XX PR 22-JUN-1999; 99US-0140650.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 20-JUL-1999; 99US-0144758.

XX PR 01-SEP-1999; 99MO-US20111.

XX PR 08-SEP-1999; 99MO-US20594.

XX PR 29-OCT-1999; 99US-0162506.

XX PR 30-NOV-1999; 99MO-US28313.

XX PR 01-DEC-1999; 99MO-US28634.

XX PR 02-DEC-1999; 99MO-US28551.

XX PR 16-DEC-1999; 99MO-US30095.

XX PR 20-DEC-1999; 99MO-US30999.

XX PR 06-JAN-2000; 2000MO-US00376.

XX PR 11-FEB-2000; 2000MO-US00341.

XX PR 18-FEB-2000; 2000MO-US04342.

XX PR 02-MAR-2000; 2000MO-US05841.

XX PR 03-MAR-2000; 2000US-0187202.

XX PR 10-MAR-2000; 2000MO-US06315.

XX PR 15-MAR-2000; 2000MO-US06884.

XX PR 30-MAR-2000; 2000MO-US08439.

XX PR 17-MAY-2000; 2000MO-US13705.

XX PA (GETH ) GENENTECH INC.

XX PI Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W, Kabakoff RC;

XX PI Shelton DL, Smith V, Watanabe CK, Wood WI;

XX DR WPI: 2001-016509/02.

XX DR N-PSDB; C91560.

PS Twenty eight nucleic acids encoding PRO polypeptides which are useful  
PT for treating various tumors, e.g. breast cancer, and other  
PT inflammatory, angiogenic and immunological disorders -

PS Claim 31; Fig 16; 188pp; English.

CC The present sequence is one of twenty eight novel PRO polypeptides. The  
CC PRO polypeptides and their agonists, including antibodies, peptides, and  
CC small molecule agonists, may be used to treat various tumors, e.g.,  
CC cancers such as breast cancer, ovarian cancer, renal cancer, colorectal  
CC cancer, uterine cancer, prostate cancer, lung cancer, bladder cancer,  
CC central nervous system cancer, melanoma or leukemia. They are also  
CC useful for treating other disorders such as neuronal, glial, astrocytal,  
CC hypothalamic and other glandular, macrophagal, epithelial, stromal and  
CC blastocoealic disorders, and inflammatory, angiogenic and immunological  
CC disorders.

SQ Sequence 346 AA;

Query Match 74.5%; Score 35; DB 22; Length 346;  
Best Local Similarity 100.0%; Pred. NO. 2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGCVD 7

DB 191 rgcvqd 196

## RESULT 12

R58666  
ID R58666 standard; Protein: 448 AA.

XX AC R58666;

XX DT 12-MAY-1995 (first entry)

XX DE Rat PACAP receptor type 1A mature protein.

XX KW Bovine; rat; pituitary adenylate cyclase activating polypeptide; PACAP;  
KW adenylate cyclase; receptor; type 1A; signal sequence;  
KW hydrophobic cluster; transmembrane; human; ACAP receptor; pituitary;  
KW gland-derived; diagnosis; neuropathy; Alzheimer's disease; gene therapy;  
KW screening; assay; neuropathy.

XX OS Rattus rattus.

XX PN EP618291-A.

XX PD 05-OCT-1994.

XX PF 24-FEB-1994; 94EP-0102757.

XX PR 26-FEB-1993; 93JP-0038755.

XX PR 05-APR-1993; 93JP-0078290.

XX PR 27-APR-1993; 93JP-0100669.

XX PR 17-MAY-1993; 93JP-0114446.

XX PR 24-JUN-1993; 93JP-0153963.

XX PR 10-NOV-1993; 93JP-0281413.

XX PR 27-DEC-1993; 93JP-0333175.

XX PA (TAKE ) TAKEDA CHEM IND LTD.

XX PI Habata Y, Hosoya M, Ishibashi Y, Kitada C, Masuda Y;

XX PI Miyamoto Y, Ogi K, Ohnaki T, Onda H, Shimamoto N;

XX DR WPI: 1994-304460/38.

XX PS Claim 3; Page 58-60; 164pp; English.

CC This is the amino acid sequence of the mature protein lacking the signal  
CC peptide of rat pituitary adenylate cyclase activating polypeptide  
CC (PACAP) receptor type 1A. The entire gene encodes a protein of 467 amino  
CC acids (53.2 kDa) including a 19 amino acid signal sequence for passing  
CC through the membrane. Type 1A and 1B PACAP receptor gene sequence differ  
CC at nucleotide 1119-20. In type 1B, there is an addition of 84 bp between  
CC these nucleotides. This produces a protein (R58658) with 28 amino acids  
CC extra than the type 1A protein. Hydrophobic measurements have deduced  
CC that the protein contains 7 hydrophobic clusters considered to be  
CC transmembrane domains in tandem. The gene for the rat PACAP receptor was  
CC cloned by a) constructing primers (Q72197-8) based on regions of high  
CC homology with the rat vasoactive intestinal peptide (VIP), b) amplifying  
CC the region of homology and c) using the resulting PCR product to screen a  
CC rat brain cDNA library to obtain a cDNA clone of the PACAP receptor. The  
CC plasmid containing the type 1A gene was called pRACAPR46-5. The entire  
CC DNA sequence including 5' and 3' regions is given in Q72205. The mature

CC sequence differs from the precursor by lacking the signal sequence. The  
CC PACAP receptor or fragments thereof may be used for the diagnosis of  
CC neuropathic such as Alzheimer's disease. The DNA encoding the PACAP  
CC receptor can be used for gene therapy. Compounds isolated using the  
CC screening assay can be used for treating neuropathy.

SO Sequence 448 AA;

Query Match 72.3%; Score 34; DB 15; Length 448;  
Best Local Similarity 62.5%; Pred. No. 3.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8  
|||:|  
Db 95 grncdedg 102

RESULT 13  
R58656  
ID R58656 standard; Protein; 467 AA.  
XX  
AC R58656;  
XX  
DT 11-MAY-1995 (first entry)  
XX  
DE Rat PACAP receptor type 1A protein.  
XX  
KW Bovine; rat; pituitary adenylate cyclase activating polypeptide; PACAP;  
KW adenylate cyclase; receptor; type 1A; signal sequence;  
KW hydrophobic cluster; transmembrane; human; ACAP receptor; pituitary;  
KW gland-derived; diagnosis; neuropathy; Alzheimer's disease; gene therapy;  
KW screening; assay; neuropathy.

OS Rattus rattus.  
XX  
FH Key Location/Qualifiers  
FT sig\_peptide 1..19 /\*tag- a  
FT mat\_peptide 20..467 /\*tag- b  
FT  
XX  
PN EP618291-A.  
XX  
PD 05-OCT-1994.  
XX  
PF 24-FEB-1994; 94EP-0102757.  
XX  
PR 26-FEB-1993; 93JP-0038755.  
PR 05-APR-1993; 93JP-0078290.  
PR 27-APR-1993; 93JP-0100669.  
PR 17-MAY-1993; 93JP-0114446.  
PR 24-JUN-1993; 93JP-0153963.  
PR 10-NOV-1993; 93JP-0281413.  
PR 27-DEC-1993; 93JP-0333175.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Habata Y, Hosoya M, Ishibashi Y, Kitada C, Masuda Y;  
PI Miyamoto Y, Ogi K, Ohtaki T, Onda H, Shimamoto N;  
XX  
DR WPI: 1994-304460/38.  
DR N-PSDB: 072189.  
XX  
PT New pituitary adenylate cyclase activating polypeptide receptor  
PT proteins - used to develop prods. for use in the diagnosis and  
PT treatment of neuropathy such as Alzheimer's disease  
XX  
PS Claim 9; Page 80; 164pp; English.  
XX  
CC This is the amino acid sequence deduced from the coding region of the rat  
CC pituitary adenylate cyclase activating polypeptide (PACAP) receptor  
CC type 1A. The entire gene encodes a protein of 467 amino acids (53.2 kDa)

CC including a 19 amino acid signal sequence for passing through the  
CC membrane. Type 1A and 1B PACAP receptor gene sequence differ at  
CC nucleotide 1119-20. In type 1B, there is an addition of 84 bp between  
CC these nucleotides. This produces a protein (R58658) with 28 amino acids  
CC more than the type 1A protein. Hydrophobic measurements have deduced  
CC that the protein contains 7 hydrophobic clusters considered to be  
CC transmembrane domains in tandem. The gene for the rat PACAP receptor was  
CC cloned by a) constructing primers (Q72197-8) based on regions of high  
CC homology with the rat vasoactive intestinal peptide (VIP), b) amplifying  
CC the region of homology and c) using the resulting PCR product to screen a  
CC rat brain cDNA library to obtain a cDNA clone of the PACAP receptor. The  
CC plasmid containing the type 1A gene was called pRACAPR46-5. The mature  
CC sequence differs from the precursor by lacking the signal peptide. The  
CC PACAP receptor or fragments thereof may be used for the diagnosis of  
CC neuropathy such as Alzheimer's disease. The DNA encoding the PACAP  
CC receptor can be used for gene therapy. Compounds isolated using the  
CC screening assay can be used for treating neuropathy.

SO Sequence 467 AA;

Query Match 72.3%; Score 34; DB 15; Length 467;  
Best Local Similarity 62.5%; Pred. No. 3.9e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8  
|||:|  
Db 114 grncdedg 121

RESULT 14  
R58667  
ID R58667 standard; Protein; 476 AA.  
XX  
AC R58667;  
XX  
DT 12-MAY-1995 (first entry)  
XX  
DE Rat PACAP receptor type 1B mature protein.  
XX  
KW Bovine; rat; pituitary adenylate cyclase activating polypeptide; PACAP;  
KW adenylate cyclase; receptor; type 1A; signal sequence;  
KW hydrophobic cluster; transmembrane; human; ACAP receptor; pituitary;  
KW gland-derived; diagnosis; neuropathy; Alzheimer's disease; gene therapy;  
KW screening; assay; neuropathy.

OS Rattus rattus.  
XX  
PN EP618291-A.  
XX  
PD 05-OCT-1994.  
XX  
PF 24-FEB-1994; 94EP-0102757.  
XX  
PR 26-FEB-1993; 93JP-0038755.  
PR 05-APR-1993; 93JP-0078290.  
PR 27-APR-1993; 93JP-0100669.  
PR 17-MAY-1993; 93JP-0114446.  
PR 24-JUN-1993; 93JP-0153963.  
PR 10-NOV-1993; 93JP-0281413.  
PR 27-DEC-1993; 93JP-0333175.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Habata Y, Hosoya M, Ishibashi Y, Kitada C, Masuda Y;  
PI Miyamoto Y, Ogi K, Ohtaki T, Onda H, Shimamoto N;  
XX  
DR WPI: 1994-304460/38.  
DR  
XX  
PT New pituitary adenylate cyclase activating polypeptide receptor  
PT proteins - used to develop prods. for use in the diagnosis and  
PT treatment of neuropathy such as Alzheimer's disease  
XX

PS Claim 3; Page 61-63; 164pp; English.

XX This is the amino acid sequence of the mature protein lacking the signal  
 CC peptide of rat pituitary adenylate cyclase activating polypeptide  
 CC (PACAP) receptor type 1B. The entire gene encodes a protein of 495 amino  
 CC acids (56.4 kDa) including a 19 amino acid signal sequence for passing  
 CC through the membrane. Type 1A and 1B PACAP receptor gene sequence differ  
 CC at nucleotide 1119-20. In type 1B, there is an addition of 84 bp between  
 CC these nucleotides. This produces a protein with 28 amino acids more than  
 CC the Type 1A protein (R58656). Hydrophobic measurements have deduced  
 CC that the protein contains 7 hydrophobic clusters considered to be  
 CC transmembrane domains in tandem. The gene for the rat PACAP receptor was  
 CC cloned by a) constructing primers (Q72197-8) based on regions of high  
 CC homology with the rat vasoactive intestinal peptide (VIP), b) amplifying  
 CC the region of homology and c) using the resulting PCR product to screen a  
 CC rat brain cDNA library to obtain a cDNA clone of the PACAP receptor. The  
 CC plasmid containing the type 1A gene was called pPACAPR46-5. The entire  
 CC DNA sequence including 5' and 3' regions is given in Q72205. The mature  
 CC sequence differs from the precursor by lacking the signal sequence. The  
 CC PACAP receptor or fragments thereof may be used for the diagnosis of  
 CC neuropathy such as Alzheimer's disease. The DNA encoding the PACAP  
 CC receptor can be used for gene therapy. Compounds isolated using the  
 CC screening assay can be used for treating neuropathy.

SQ Sequence 476 AA;

Query Match 72.3%; Score 34; DB 15; Length 476;

Best Local Similarity 62.5%; Pred. No. 3.9e+02; Mismatches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8

DB 95 grncdtdg 102

RESULT 15

ID R58658 standard; Protein; 495 AA.

AC R58658;

DT 11-MAY-1995 (first entry)

DE Rat PACAP receptor type 1B protein.

KW Bovine; rat; pituitary adenylate cyclase activating polypeptide; PACAP;

KW adenylate cyclase; receptor; type 1A; signal sequence;

KW hydrophobic cluster; transmembrane; human; ACAP receptor; pituitary;

KW gland-derived; diagnosis; neuropathy; Alzheimer's disease; gene therapy;

XX screening; assay; neuropathy.

OS Rattus rattus.

FX Key

FT sig\_peptide

FT mat\_peptide

FT /tag= a

FT /tag= b

XX EP618291-A.

XX 05-OCT-1994.

XX 24-FEB-1994; 94EP-0102757.

XX 26-FEB-1993; 93JP-0038755.

XX 05-APR-1993; 93JP-0078290.

XX 27-APR-1993; 93JP-0100669.

XX 17-MAY-1993; 93JP-0114446.

XX 24-JUN-1993; 93JP-0153963.

XX 10-NOV-1993; 93JP-0281413.

XX 27-DEC-1993; 93JP-0333175.

XX (TAKE ) TAKEDA CHEM IND LTD.

PI Habata Y, Hosoya M, Ishibashi Y, Kitada C, Masuda Y;

PI Miyamoto Y, Ogi K, Ohtaki T, Onda H, Shimamoto N;

DR MPI: 1994-304460/38.

DR N-PSDB; Q72191.

PT New pituitary adenylate cyclase activating polypeptide receptor

PT proteins - used to develop prods. for use in the diagnosis and

PT treatment of neuropathy such as Alzheimer's disease

XX Claim 9; Page 81; 164pp; English.

XX This is the amino acid sequence deduced from the coding region of the rat  
 CC pituitary adenylate cyclase activating polypeptide (PACAP) receptor  
 CC type 1B. The entire gene encodes a protein of 495 amino acids (56.4 kDa)  
 CC including a 19 amino acid signal sequence for passing through the  
 CC membrane. Type 1A and 1B PACAP receptor gene sequence differ at  
 CC nucleotide 1119-20. In type 1B, there is an addition of 84 bp between  
 CC these nucleotides. This produces a protein with 28 amino acids more than  
 CC the Type 1A protein (R58656). Hydrophobic measurements have deduced  
 CC that the protein contains 7 hydrophobic clusters considered to be  
 CC transmembrane domains in tandem. The gene for the rat PACAP receptor was  
 CC cloned by a) constructing primers (Q72197-8) based on regions of high  
 CC homology with the rat vasoactive intestinal peptide (VIP), b) amplifying  
 CC the region of homology and c) using the resulting PCR product to screen a  
 CC rat brain cDNA library to obtain a cDNA clone of the PACAP receptor. The  
 CC plasmid containing the type 1A gene was called pPACAPR46-5. The mature  
 CC sequence differs from the precursor by lacking the signal peptide. The  
 CC PACAP receptor or fragments thereof may be used for the diagnosis of  
 CC neuropathy such as Alzheimer's disease. The DNA encoding the PACAP  
 CC receptor can be used for gene therapy. Compounds isolated using the  
 CC screening assay can be used for treating neuropathy.

SQ Sequence 495 AA;

Query Match 72.3%; Score 34; DB 15; Length 495;

Best Local Similarity 62.5%; Pred. No. 4.1e+02; Mismatches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8

DB 114 grncdtdg 121

Search completed: June 13, 2001, 14:14:35  
 Job time: 376 sec



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:42 : Search time 130.61 Seconds

(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-10

Sequence: 1 GRGCVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR.67:\*  
2: PIR.7:\*  
3: PIR.3:\*  
4: PIR.4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	78.7	399	2 B82197	zona occludens tox
2	37	78.7	399	2 A43864	zona occludens t
3	36	76.6	206	2 S18250	collagen alpha 1(I
4	36	76.6	814	2 T05827	hypothetical prote
5	35	74.5	188	2 H38192	chitin synthase (E
6	35	74.5	449	2 S48173	variable surface g
7	35	74.5	993	2 A59054	chitin synthase (E
8	34	72.3	162	2 A27398	allophycocyanin al
9	34	72.3	296	2 A70973	hypothetical prote
10	34	72.3	467	2 JN0616	pituitary adenylat
11	34	72.3	495	2 S36114	pituitary adenylat
12	34	72.3	495	2 S39061	pituitary adenylat
13	34	72.3	523	2 S39060	pituitary adenylat
14	34	72.3	796	2 T23238	hypothetical prote
15	34	72.3	1242	1 DJBEC1	DNA-directed DNA P
16	34	72.3	3436	2 S55659	tegment protein 6
17	33	70.2	130	2 F72586	hypothetical prote
18	33	70.2	133	2 E69338	hypothetical prote
19	33	70.2	335	2 B72053	glyceroldehyde 3-P
20	33	70.2	491	2 S56753	interferon regulat
21	33	70.2	556	2 S22634	sphingomyelin phos
22	33	70.2	617	2 T07612	cellulase (EC 3.2.
23	33	70.2	658	2 T08153	cysteine proteinas
24	33	70.2	680	1 XJBYTK	transketolase (EC
25	33	70.2	681	2 S37809	transketolase (EC
26	33	70.2	685	2 T40162	transketolase (EC
27	33	70.2	695	2 T36007	probable transketo
28	33	70.2	983	2 H72510	probable ribonucle
29	33	70.2	2240	2 T37057	probable multi-dom

30	32	68.1	136	2 E72759	hypothetical prote
31	32	68.1	142	2 T36312	hypothetical prote
32	32	68.1	224	2 C72010	conserved hypothet
33	32	68.1	302	2 G70614	hypothetical prote
34	32	68.1	335	2 T41426	hypothetical wtf8
35	32	68.1	337	2 T25210	hypothetical prote
36	32	68.1	402	2 T41253	hypothetical wtf5
37	32	68.1	411	2 S41648	translational elonga
38	32	68.1	418	2 T41027	hypothetical wtf5
39	32	68.1	540	2 T08662	probable signaling
40	32	68.1	648	2 T23864	hypothetical prote
41	32	68.1	1353	1 JH0675	restrictin precurs
42	32	68.1	1447	2 T42628	neuronal apoptosis
43	32	68.1	1562	2 T43022	ATP-binding multid
44	31	66.0	133	2 T04081	probable ribosomal
45	31	66.0	196	2 F71525	hypothetical prote

## ALIGNMENTS

RESULT 1  
B82197  
zona occludens toxin VC1458 [Imported] - Vibrio cholerae (group O1 strain N16961)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 15-Sep-2000  
C:Accession: B82197  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406: 477-483, 2000  
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: B82197  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-399 <HEI>  
A:Cross-references: GB:AE004224; GB:AE003852; NID:g9655952; PIDN:AMF94615.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1458  
A:Map position: 1

Query Match 78.7%; Score 37; DB 2; Length 399;  
Best Local Similarity 87.5%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GRGCVQDG 8  
DB 291 GRGCVQDG 298

RESULT 2  
A43864  
zona occludens toxin - Vibrio cholerae  
C:Species: Vibrio cholerae  
C:Date: 10-Mar-1993 #sequence-revision 18-Nov-1994 #text-change 08-Oct-1999  
C:Accession: A43864  
R:Bandry, B.; Fasano, A.; Kelley, J.; Kaper, J.B.  
Infect. Immun. 60: 428-434, 1992  
A:Title: Cloning of a gene (zot) encoding a new toxin produced by Vibrio cholerae.  
A:Reference number: A43864; MUID:92113300  
A:Accession: A43864  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-399 <BAU>  
A:Cross-references: GB:M83563; NID:g155314; PIDN:AAA27582.1; PID:g155315  
A:Note: sequence extracted from NCBI backbone (NCBIN:77488, NCBIIP:77491)

Query Match 78.7%; Score 37; DB 2; Length 399;  
Best Local Similarity 87.5%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GRGCVQDG 8  
 |||||  
 Db 291 GRGCVQDG 298

## RESULT 3

collagen alpha 1(II) chain precursor - chicken (fragment)  
 C:Species: Gallus gallus (chicken)  
 C>Date: 22-Nov-1993 #sequence\_revision 26-Jul-1996 #text\_change 13-Aug-1999  
 C:Accession: S18250  
 R:NaN, H.D.; Upholt, W.B. 23446-23452, 1991  
 J. Biol. Chem. 266, 23446-23452, 1991  
 A:Title: Type II collagen mRNA containing an alternatively spliced exon predominates in  
 A:Reference number: S18250; MUID:92078225  
 A:Accession: S18250  
 A:Molecule type: mRNA  
 A:Residues: 1-206 <NAH>  
 A:Cross-references: EMBL:M74435; NID:g211635; PIDN:AAA48714.1; PID:g211636  
 C:Gene: COL2A1  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-184/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:36-95/Domain: von Willebrand factor type C repeat homology <WMC>  
 F:185-206/Product: collagen alpha 1(II) chain (fragment) #status predicted <MAT>

Query Match 76.6%; Score 36; DB 2; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 GCYQDG 8  
 |||||  
 Db 36 GCYQDG 41

## RESULT 4

hypothetical protein T5K18.220 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 13-Aug-1999  
 C:Accession: T05827  
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.-J.; Voet, M.; Robben, J.; Volckaert, G.; Be  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: T05827  
 A:Accession: T05827  
 A:Molecule type: DNA  
 A:Residues: 1-814 <BEV>  
 A:Cross-references: EMBL:AL022580  
 A:Experimental source: cultivar Columbia; BAC clone T5K18  
 C:Genetics:  
 A:Map position: 4  
 A>Note: T5K18.220

Query Match 76.6%; Score 36; DB 2; Length 814;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVQD 7  
 |||||  
 Db 522 GRGCVQD 528

## RESULT 5

chitin synthase (EC 2.4.1.16) CHS1 - Aspergillus niger (fragment)  
 C:Species: Aspergillus niger  
 C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Aug-1999

C:Accession: H38192  
 R:Bowen, A.R.; Chen-Wu, J.L.; Momany, M.; Young, R.; Szaniszló, P.J.; Robbins, P.W.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 519-523, 1992  
 A:Title: Classification of fungal chitin synthases.  
 A:Reference number: A38192; MUID:92115692  
 A:Accession: H38192

A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-188 <BOW>  
 A>Note: the accession cited in the publication, M82942, appears in GenBank EMBL ent  
 A:Note: sequence extracted from NCBI Backbone (NCBIP:75830)  
 C:Superfamily: chitin synthase chsa  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 74.5%; Score 35; DB 2; Length 188;  
 Best Local Similarity 75.0%; Pred. No. 25;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
 |||||  
 Db 57 GRGCVQDG 64

## RESULT 6

variable surface glycoprotein (clone BENat 1.2) - Trypanosoma congolense  
 C:Species: Trypanosoma congolense  
 C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 09-Sep-1997  
 C:Accession: S48173; S45056  
 R:Rausch, S.; Shayan, P.; Salnikoff, J.; Reinwald, E.  
 Eur. J. Biochem. 223, 813-821, 1994  
 A:Title: Sequence determination of three variable surface glycoproteins from Trypanos  
 A:Reference number: S48172; MUID:94333385  
 A:Accession: S48173  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-449 <RAU>  
 A:Cross-references: EMBL:X79400; NID:9495151; PID:9495152  
 C:Keywords: glycoprotein

Query Match 74.5%; Score 35; DB 2; Length 449;  
 Best Local Similarity 75.0%; Pred. No. 53;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
 |||||  
 Db 207 GRGCVQDG 214

## RESULT 7

chitin synthase (EC 2.4.1.16) CHS1 - Emericella nidulans  
 N:Alternate names: chitin synthase chsc  
 C:Species: Emericella nidulans; Aspergillus nidulans  
 C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: A59054; JC2408; G38192  
 R:Horinuchi, H.  
 submitted to DDBJ, February 1999  
 A:Description: Emericella nidulans gene for chitin synthase, complete cds.  
 A:Reference number: A59054  
 A:Accession: A59054  
 A:Molecule type: DNA  
 A:Residues: 1-983 <HOR>  
 A:Cross-references: GB:AB023911; DDBJ:D8409; NID:94519180; PIDN:BAAT5501.1; PID:9451  
 A:Experimental source: strain FGSC A89  
 R:Motoyama, T.; Kojima, N.; Horinuchi, H.; Ohta, A.; Takagi, M.  
 Biosci. Biotechnol. Biochem. 58, 2254-2257, 1994  
 A:Title: Isolation of a chitin synthase gene (chsc) of Aspergillus nidulans.  
 A:Reference number: JC2408; MUID:95128032  
 A:Accession: JC2408  
 A:Molecule type: DNA

A:Residues: 1-18, 'A', 20-85, 'TNTIOYTSRPMRKHTMTLIPNHTSLQ', 'LHRAMTMTSASIDISITRMSRSFNF  
868-889, 'LGYCWMQVSRSLFVRCGFACEN' <NOT>  
A:Cross-references: DDBJ:D8409; NID:94519180  
R:Bowen, A.R.; Chen-Wu, J.L.; Momeny, M.; Young, R.; Szaniszlo, P.J.; Robbins, P.W.  
Proc. Natl. Acad. Sci. U.S.A. 89, 519-523, 1992  
A:Title: Classification of fungal chitin synthases.  
A:Reference number: A38192; MUID:92115692  
A:Accession: G38192  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 258-280, 'RS', 283-446 <BOW>  
A:Cross-references: GB:M82938; NID:g168038; PIDN:AAA3302.1; PID:g168039  
A:Note: sequence extracted from NCBI backbone (NCBI:75825)  
C:Genetics:  
A:Gene: chsc  
A:Introns: 7/1; 66/3; 611/3; 745/3  
C:Function:  
A:Description: catalyzes the alpha-1,4-glycosylation of chitin by UDP-N-acetyl-D-glucose  
C:Superfamily: chitin synthase chsa  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 74.5%; Score 35; DB 2; Length 983;  
Best Local Similarity 75.0%; Pred. No. 1,1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8  
1 1 1 1 1 1  
Db 314 GAGCVQDG 321

RESULT 8  
A27398  
Allophycocyanin alpha-B chain - *Synechococcus* sp.  
C:Species: *Synechococcus* sp.  
C:Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 30-Apr-1999  
C:Accession: A27398  
R:Steter, F.; Fuglistaller, P.; Lundell, D.J.; Glaeser, A.N.; Zuber, H.  
FEBS Lett. 217, 279-282, 1987  
A:Title: Amino acid sequences of alpha-allophycocyanin B from *Synechococcus* 6301 and Mas  
A:Reference number: A91377  
A:Accession: A27398  
A:Molecule type: protein  
A:Residues: 1-162 <SUT>  
A:Experimental source: strain 6301  
C:Superfamily: phycocyanin  
C:Keywords: chromoprotein; photosynthesis; phycocyanobilin  
F:82/Binding site: phycocyanobilin (Cys) (covalent) #status predicted

Query Match 72.3%; Score 34; DB 2; Length 162;  
Best Local Similarity 71.4%; Pred. No. 34;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGCYQDG 8  
1 1 1 1 1 1  
Db 80 RGCURDG 86

RESULT 9  
A70973  
hypothetical protein RV3378c - *Mycobacterium tuberculosis* (strain H37Rv)  
C:Species: *Mycobacterium tuberculosis*  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70973  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Randeram, M.A.; Rogers, J.; Ruter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: A70973

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-296 <COL>  
A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15763.1; PID:e120  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3378c

Query Match 72.3%; Score 34; DB 2; Length 296;  
Best Local Similarity 75.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8  
1 1 1 1 1 1  
Db 283 GVCVHDG 290

RESULT 10  
JN0616  
pituitary adenylate cyclase-activating polypeptide receptor 46-5 precursor - rat  
N:Alternate names: PACAP receptor  
C:Species: *Rattus norvegicus* (Norway rat)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 08-Oct-1999  
C:Accession: JN0616; S36768  
R:Hosoya, M.; Onda, H.; Ogi, K.; Masuda, Y.; Miyamoto, Y.; Ohtaki, T.; Okazaki, H.; A  
Biochem. Biophys. Res. Commun. 194, 133-143, 1993  
A:Title: Molecular cloning and functional expression of rat cDNAs encoding the recept  
A:Reference number: JN0616; MUID:93326107  
A:Accession: JN0616  
A:Molecule type: mRNA  
A:Residues: 1-467 <HOS>  
A:Experimental source: brain  
R:Spengler, D.; Waerber, C.; Pantaloni, C.; Holsboer, F.; Bockaert, J.; Seeburg, P.H.;  
Nature 365, 170-175, 1993  
A:Title: Differential signal transduction by five splice variants of the PACAP recept  
A:Reference number: S36768; MUID:93382505  
A:Accession: S36768  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-467 <SPE>  
A:Cross-references: EMBL:223279; NID:g404252; PIDN:CAA80817.1; PID:g404253  
C:Superfamily: glucagon receptor  
C:Keywords: alternative splicing; glycoprotein; receptor  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-467/Product: pituitary adenylate cyclase-activating polypeptide receptor 46-5 #s  
F:47,59,116,299,342,374/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 72.3%; Score 34; DB 2; Length 467;  
Best Local Similarity 62.5%; Pred. No. 84;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8  
1 1 1 1 1 1  
Db 114 GRNCRDGD 121

RESULT 11  
S36114  
pituitary adenylate cyclase-activating polypeptide type I receptor precursor - rat  
N:Alternate names: PACAP receptor  
C:Species: *Rattus norvegicus* (Norway rat)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 01-Dec-2000  
C:Accession: S36114; JN0617; S39062; S39063; PN0609; I58147; A48204  
R:Morrow, J.A.; Lutz, E.M.; West, K.M.; Fink, G.; Harnar, A.J.  
FEBS Lett. 329, 99-105, 1993  
A:Title: Molecular cloning and expression of a cDNA encoding a receptor for pituitary  
A:Reference number: S36114; MUID:9335075  
A:Accession: S36114  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-495 <MOR>

A:Cross-references: EMBL:Z23282; NID:9397520; PIDN:CAA80820.1; PID:9397521  
 R:Hosoya, M.; Onda, H.; Ogi, K.; Masuda, Y.; Miyamoto, Y.; Ohtaki, T.; Okazaki, H.; Arim  
 Biochem. Biophys. Res. Commun. 194, 133-143, 1993  
 A:Title: Molecular cloning and functional expression of rat cDNAs encoding the receptor  
 A:Reference number: JN0616; MUID:93326107  
 A:Accession: JN0617  
 A:Molecule type: mRNA  
 A:Residues: 1-495 <RCS>  
 A:Experimental source: brain  
 R:Spengler, D.; Waerber, C.; Pantaloni, C.; Holsboer, F.; Bockaert, J.; Seeburg, P.H.; Jo  
 Nature 365, 170-175, 1993  
 A:Title: Differential signal transduction by five splice variants of the PACAP receptor.  
 A:Reference number: S36768; MUID:93382505  
 A:Accession: S39062  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-495 <SPE>  
 A:Cross-references: EMBL:Z23274; NID:9404220; PIDN:CAA80812.1; PID:9404221  
 A:Accession: S39063  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-348 350-495 <SP2>  
 A:Cross-references: EMBL:Z23275; NID:9404197; PIDN:CAA80813.1; PID:9404198  
 R:Spengler, D.; Waerber, C.; Pantaloni, C.; Holsboer, F.; Bockaert, J.; Seeburg, P.H.; Jo  
 Biochem. Biophys. Res. Commun. 195, 881-888, 1993  
 A:Title: Cloning of a splice variant of the pituitary adenylyl cyclase-activating poly  
 A:Reference number: PNO608; MUID:93384616  
 A:Accession: PNO608  
 A:Molecule type: mRNA  
 A:Residues: 78-495 <SVO>  
 A:Accession: PNO609  
 A:Molecule type: protein  
 A:Residues: 349-376 <SV2>  
 R:Hashimoto, H.; Ishihara, T.; Shigemoto, R.; Morl, K.; Nagata, S.  
 Neuron 11, 333-342, 1993  
 A:Title: Molecular cloning and tissue distribution of a receptor for pituitary adenylyl  
 A:Reference number: I58147; MUID:93357025  
 A:Accession: I58147  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-495 <RES>  
 A:Cross-references: GB:D16465; NID:9440381; PIDN:BA03932.1; PID:9457661  
 R:Pisegna, J.R.; Wank, S.A.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 6345-6349, 1993  
 A:Title: Molecular cloning and functional expression of the pituitary adenylyl cyclase-  
 A:Reference number: A48204; MUID:93317678  
 A:Accession: A48204  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-420, 'U', 422-495 <PIS>  
 A:Cross-references: GB:L16680; NID:9347941; PIDN:AAA1192.1; PID:9347942  
 C:Comment: This protein stimulates both adenylyl cyclase and phospholipase C and dually  
 C:Superfamily: glucagon receptor  
 C:Keywords: alternative splicing; anterior pituitary; G protein-coupled receptor; glyco  
 F:1-495/Product: pituitary adenylyl cyclase-activating polypeptide receptor long form #  
 F:1-377,376-495/Product: pituitary adenylyl cyclase-activating polypeptide receptor shc  
 F:1-19/Domains: signal sequence #status predicted <SIG>  
 F:20-495/Product: pituitary adenylyl cyclase-activating polypeptide type I receptor #st  
 F:155-177/Domains: transmembrane #status predicted <TM1>  
 F:186-204/Domains: transmembrane #status predicted <TM2>  
 F:227-252/Domains: transmembrane #status predicted <TM3>  
 F:268-290/Domains: transmembrane #status predicted <TM4>  
 F:308-331/Domains: transmembrane #status predicted <TM5>  
 F:378-398/Domains: transmembrane #status predicted <TM6>  
 F:413-433/Domains: transmembrane #status predicted <TM7>  
 F:47-59,116,299,342,402/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:365,444/Binding site: phosphate (Ser) (covalent) #status predicted

QY 1 GRGCVODG 8  
 ||| :||  
 Db 114 GRNCTEDG 121  
 RESULT 12  
 S39061  
 pituitary adenylyl cyclase activating-peptide receptor form 3 - rat  
 M:Alternate names: PACAP receptor  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999  
 C:Accession: S39061  
 R:Spengler, D.; Waerber, C.; Pantaloni, C.; Holsboer, F.; Bockaert, J.; Seeburg, P.H.;  
 Nature 365, 170-175, 1993  
 A:Title: Differential signal transduction by five splice variants of the PACAP recept  
 A:Reference number: S36768; MUID:93382505  
 A:Accession: S39061  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-495 <SPE>  
 A:Cross-references: EMBL:Z23273; NID:9404210; PIDN:CAA80811.1; PID:9404211  
 C:Superfamily: glucagon receptor

Query Match 72.3%; Score 34; DB 2; Length 495;  
 Best Local Similarity 62.5%; Pred. No. 88;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GRGCVODG 8  
 ||| :||  
 Db 114 GRNCTEDG 121

RESULT 13  
 S39060  
 pituitary adenylyl cyclase activating-peptide receptor form 2 - rat  
 M:Alternate names: PACAP receptor  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Nov-1999  
 C:Accession: S39060  
 R:Spengler, D.; Waerber, C.; Pantaloni, C.; Holsboer, F.; Bockaert, J.; Seeburg, P.H.;  
 Nature 365, 170-175, 1993  
 A:Title: Differential signal transduction by five splice variants of the PACAP recept  
 A:Reference number: S36768; MUID:93382505  
 A:Accession: S39060  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-523 <SP2>  
 A:Cross-references: EMBL:Z23272; NID:9404195; PIDN:CAA80810.1; PID:9404196  
 C:Superfamily: glucagon receptor

Query Match 72.3%; Score 34; DB 2; Length 523;  
 Best Local Similarity 62.5%; Pred. No. 93;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GRGCVODG 8  
 ||| :||  
 Db 114 GRNCTEDG 121

RESULT 14  
 T23238  
 hypothetical protein T05H10.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T23238; T24560  
 R:Lightning, J.  
 submitted to the EMBL Data Library, January 1995  
 A:Reference number: Z19713  
 A:Accession: T23238  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-796 <WIL>  
 A:Cross-references: EMBL:Z47811; PIDN:CAA87788.1; GSPDB:GN00020; CESP:T05H10.7  
 A:Experimental source: clone K02C4  
 R:Thomas, K.  
 submitted to the EMBL Data Library, January 1995  
 A:Reference number: Z19907  
 A:Accession: T24560  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-796 <W12>  
 A:Cross-references: EMBL:Z47812; PIDN:CAA87796.1; GSPDB:GN00020; CESP:T05H10.7  
 A:Experimental source: clone T05H10  
 C:Genetics:  
 A:Gene: CESP:T05H10.7  
 A:Map position: 2  
 A:Introns: 77/1; 128/2; 163/1; 249/3; 559/3; 561/3; 648/3; 776/3

Query Match 72.3%; Score 34; DB 2; Length 796;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8  
 |||||  
 Db 198 GRGCVSDG 205

## RESULT 15

DJBEC1  
 DNA-directed DNA polymerase (EC 2.7.7.7) - human cytomegalovirus (strain AD169)  
 N:Alternate names: HPLF2 protein; UL54 protein  
 C:Species: human cytomegalovirus, human herpesvirus 5  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Dec-1990 #text\_change 11-Jun-1999  
 C:Accession: S09817; A25983  
 R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; M.; Barrell, B.G.  
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
 A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
 A:Reference number: S09749; MUID:90269039  
 A:Accession: S09817  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1242 <CHE>  
 A:Cross-references: EMBL:X17403; NID:959591; PIDN:CAA35413.1; PID:q1780832  
 A:Note: possible protein-coding frames are given  
 A:Note: the DNA sequence was submitted to EMBL, December 1989, in computer-readable form  
 R:Kuzariades, T.; Bankier, A.T.; Satchwell, S.C.; Weston, K.; Tomlinson, P.; Barrell, B.  
 J. Virol. 61, 125-133, 1987  
 A:Title: Sequence and transcription analysis of the human cytomegalovirus DNA polymerase  
 A:Reference number: A25983; MUID:87061230  
 A:Accession: A25983  
 A:Molecule type: DNA  
 A:Residues: 1-1242 <ROU>  
 A:Cross-references: GB:M14709; NID:g330640; PIDN:AAA45988.1; PID:g330642  
 C:Genetics:  
 A:Map position: unique long region (U-L)  
 C:Superfamily: herpesvirus DNA-directed DNA polymerase  
 C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 72.3%; Score 34; DB 1; Length 1242;  
 Best Local Similarity 62.5%; Pred. No. 2e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8  
 |||||  
 Db 347 GRGCTSEG 354

Search completed: June 13, 2001, 14:10:43  
 Job time: 144 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:21:43 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825A-10  
Perfect score: 47  
Sequence: 1 GRGCVQDG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	78.7	399	1 ZOT_VIRBCH	P38442 vibrio chol
2	35	74.5	189	1 CHS1_ASPNG	P30581 aspergillus
3	35	74.5	189	1 CHS1_EMEMI	P30583 emericella
4	35	74.5	1120	1 DPOL_RCMVM	O85428 rat cytomeg
5	34	72.3	162	1 PHAC-STNPE	P11390 synechococ
6	34	72.3	496	1 PACR_MOUSE	P70205 mus musculu
7	34	72.3	523	1 PACR_MOUSE	P70205 mus musculu
8	34	72.3	796	1 YRS7_GAEEL	Q10003 caenorhabdi
9	34	72.3	1242	1 DPOL_HCMVA	P08546 human cytom
10	33	70.2	335	1 G3P_CHLPN	O82740 chlamydia p
11	33	70.2	343	1 GLGD_BACST	O08327 bacillus st
12	33	70.2	491	1 TRF3_CHICK	O30643 gallus gall
13	33	70.2	556	1 PHL_LEPIN	P17627 leptospira
14	33	70.2	621	1 YC92_GAEEL	P51115 caenorhabdi
15	33	70.2	679	1 TKTL1_KILUA	Q12630 kuuyveromyc
16	33	70.2	679	1 TKTL1_YEAST	P33254 saccharomyc
17	33	70.2	681	1 TKTL1_YEAST	P33254 saccharomyc
18	33	70.2	681	1 TKTL1_YEAST	P33254 saccharomyc
19	33	70.2	695	1 TKT_RHINE	P36900 thizobium m
20	32	68.1	511	1 EF1G_TRYCR	P44715 trypanosoma
21	32	68.1	1171	1 C4A2_DROME	O49588 drosophila
22	32	68.1	1171	1 DPOL_HSVT1	O49588 drosophila
23	32	68.1	1171	1 DPOL_HSVT2	O49588 drosophila
24	31	66.0	133	1 BIRB_MOUSE	O49588 drosophila
25	31	66.0	220	1 RS26_ORYSA	P49216 mus musculu
26	31	66.0	247	1 CAH_METTE	Q10790 oryza sativ
27	31	66.0	340	1 ASPG_FLAME	O47898 mycobacteri
28	31	66.0	340	1 HUR_STRAU	O00923 streptococ
29	31	66.0	366	1 AROC_NEIMA	O01161 neisseria m
30	31	66.0	366	1 AROC_NEIMB	O01161 neisseria m
31	31	66.0	440	1 SCRC_HUMAN	Q47872 homo sapien
32	31	66.0	445	1 SCRC_RABIT	O46502 corycolagus
33	31	66.0	447	1 AROC_CORSE	P27793 corycolagus

## ALIGNMENTS

RESULT 1				ALIGNMENTS
ZOT_VIRBCH	STANDARD:	PRT:	399 AA.	
ID_ZOT_VIRBCH	P38442: 09L706: 09R3V6:			
AC	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	ZONA OCCUDENS TOXIN (ZONULAR OCCUDENS TOXIN).			
GN	ZOT OR VC1458.			
OS	Vibrio cholerae.			
OC	Bacteria: Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CLASSICAL INABA 569B;			
RC	MEDLINE=92112300; PubMed=1730472;			
RX	Baudry B., Pasano A., Kettley J., Kaper J.B.;			
RA	"Cloning of a gene (zot) encoding a new toxin produced by Vibrio			
RT	cholerae.";			
RL	Intect. Immun. 60:428-434(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-KNH002;			
RC	Shin H.J., Park Y.C., Kim Y.C.;			
RA	"Cloning and nucleotide sequence analysis of the virulence gene			
RT	cassette from Vibrio cholerae KNH002 isolated in Korea.";			
RL	Misamunhag Hoiji 35:205-210(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-O139-TOR OGAWA;			
RC	Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.;			
RA	"Cloning and Expression of zot gene from Vibrio cholerae.";			
RT	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RL	[4]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-EL TOR 86015 / SEROTYPE O1;			
RC	Kan B., Liu Y.Q., Qi G.M., Gao S.Y.;			
RA	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.			
RL	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-EL TOR N1961 / SEROTYPE O1;			
RC	MEDLINE=20406833; PubMed=10952301;			
RA	Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,			
RA	Dooson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,			
RA	Enmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,			
RA	McDonald L., Uterback T., Fleischman R.D., Nieman W.C., White O.,			
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RT	cholerae.";			
RL	Nature 406:477-483(2000).			
RN	[6]			
RP	CHARACTERIZATION.			
RC	MEDLINE=91271365; PubMed=2052603;			
RC	Pasano A., Baudry B., Pumplin D.W., Wasserman S.S., Tall B.D.,			
RA				

RA Kelley J.M., Kaper J.B.:  
 RT "Vibrio cholerae produces a second enterotoxin, which affects  
 RL intestinal tight junctions.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 88:5242-5246(1991).  
 CC -1- FUNCTION: INCREASES THE PERMEABILITY OF THE SMALL INTESTINE MUCOSA  
 CC BY AFFECTING THE STRUCTURE OF INTERCELLULAR TIGHT JUNCTIONS  
 CC (ZONULA OCCUDENS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M83563; AAA27582.1; -  
 CC DR EMBL: AF175708; AAD51358.1; -  
 CC DR EMBL: AF123049; AAD26854.1; -  
 CC DR EMBL: AF220606; AAF29547.1; -  
 CC DR EMBL: AE004224; AAF94615.1; -  
 CC PIR: A43864; A43864.  
 CC TIGR: VC1458; -  
 CC Enterotoxin; Toxin.  
 CC KW VARIANT 45 45 M -> I (IN STRAIN 569B).  
 CC FT VARIANT 100 100 V -> A (IN STRAINS 569B AND 86015).  
 CC FT VARIANT 272 272 V -> A (IN STRAIN 569B).  
 CC FT VARIANT 281 281 V -> A (IN STRAIN 569B).  
 CC FT VARIANT 349 349 A -> S (IN STRAIN 86015).  
 CC FT VARIANT 381 381 K -> R (IN STRAIN 86015).  
 CC FT CONFLICT 386 399 IKTENDKKGINSIF -> VKKEKESITKISFL (IN REF.  
 CC FT 4).  
 CC SQ SEQUENCE 399 AA; 44903 MW; 3C742AB58176774 CRC64;  
 CC -----  
 CC Query Match 78.7%; Score 37; DB 1; Length 399;  
 CC Best Local Similarity 87.5%; Pred. No. 9.9;  
 CC Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 CC -----  
 CC QY 1 GRGCVODG 8  
 CC DB 291 GRGCVODG 298  
 CC -----  
 CC RESULT 2  
 CC CHSL\_ASPNG STANDARD; PRT; 189 AA.  
 CC AC P30581;  
 CC DT 01-APR-1993 (Rel. 25, Created)  
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE CHITIN SYNTHASE 1 (EC 2.4.1.16) (CHITIN-UDP ACETYL-GLUCOSAMINYL  
 CC TRANSFERASE 1) (CLASS-I CHITIN SYNTHASE 1) (FRAGMENT).  
 CC GN CHSL.  
 CC OS Aspergillus niger.  
 CC CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC CC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 CC OX NCBI\_TaxID=5061;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=92115692; PubMed=1731323;  
 CC RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,  
 CC RA Robbins P.W.;  
 CC RT "Classification of fungal chitin synthases.";  
 CC CC Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).  
 CC -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-  
 CC BETA-D-GLUCOSAMINYL)](N) -> UDP + [1,4-(N-ACETYL-BETA-D-  
 CC GLUCOSAMINYL)](N+1).  
 CC CC -1- SUBCELLULAR LOCATION: PLASMA MEMBRANE-BOUND.  
 CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS  
 CC 1.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M82940; AAA3304.1; ALT\_SEQ.  
 CC DR PIR: H38192; H38192.  
 CC DR InterPro: IPR002923; -  
 CC DR Pfam: PF01644; Chitin\_synth. 1.  
 CC KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;  
 CC KW Multigene family.  
 CC FT NON\_TER 1 1  
 CC FT NON\_TER 189 189  
 CC SQ SEQUENCE 189 AA; 21062 MW; D64138A337DC729 CRC64;  
 CC -----  
 CC Query Match 74.5%; Score 35; DB 1; Length 189;  
 CC Best Local Similarity 75.0%; Pred. No. 12;  
 CC Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 CC -----  
 CC QY 1 GRGCVODG 8  
 CC DB 57 GRGCVODG 64  
 CC -----  
 CC RESULT 3  
 CC CHSL\_EMENT STANDARD; PRT; 189 AA.  
 CC ID CHSL\_EMENT  
 CC AC P30583;  
 CC DT 01-APR-1993 (Rel. 25, Created)  
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE CHITIN SYNTHASE 1 (EC 2.4.1.16) (CHITIN-UDP ACETYL-GLUCOSAMINYL  
 CC TRANSFERASE 1) (CLASS-I CHITIN SYNTHASE 1) (FRAGMENT).  
 CC GN CHSL.  
 CC OS Emericella nidulans (Aspergillus nidulans).  
 CC CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC CC Eurotiales; Trichocomaceae; Emericella.  
 CC OX NCBI\_TaxID=5072;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=92115692; PubMed=1731323;  
 CC RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,  
 CC RA Robbins P.W.;  
 CC RT "Classification of fungal chitin synthases.";  
 CC CC Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).  
 CC -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-  
 CC BETA-D-GLUCOSAMINYL)](N) -> UDP + [1,4-(N-ACETYL-BETA-D-  
 CC GLUCOSAMINYL)](N+1).  
 CC CC -1- SUBCELLULAR LOCATION: PLASMA MEMBRANE-BOUND.  
 CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS  
 CC 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M82938; AAA3302.1; -  
 CC DR PIR: G38192; G38192.  
 CC DR InterPro: IPR002923; -  
 CC DR Pfam: PF01644; Chitin\_synth. 1.  
 CC KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;  
 CC KW Multigene family.  
 CC FT NON\_TER 1 1  
 CC FT NON\_TER 189 189



SQ SEQUENCE 189 AA: 21244 MW: 712C54EBA02F026E CRC64:  
 Query Match 74.5%; Score 35; DB 1; Length 189;  
 Best Local Similarity 75.0%; Pred. NO. 12;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 GRCVQDG 8  
 DB 57 GRCVQDG 64  
 RESULT 4  
 DPOL\_RCMVM STANDARD: PRT: 1120 AA.  
 AC 085428;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA POLYMERASE (EC 2.7.7.7).  
 GN UL34.  
 OS Rat cytomegalovirus (strain Mastricht).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Muromegalovirus.  
 OX NCBI\_TaxID=79700;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96335691; PubMed=8757999;  
 RA Beuken E., Slobbe R., Brugeman C.A., Vink C.;  
 RT "Cloning and sequence analysis of the genes encoding DNA polymerase,  
 RT glycoprotein B, ICp18.5 and major DNA-binding protein of rat  
 RT cytomegalovirus";  
 RL J. Gen. Virol. 77:1559-1562(1996).  
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -  
 CC N RYPOHOSPHATE + DNA(N).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U50550; AAC56433.1; -  
 DR InterPro: IPR002064; -  
 DR Pfam: PF00136; DNA\_POL\_B: 1.  
 DR PRINTS: PR00106; DNAPOLB.  
 DR PROSITE: PS00116; DNA\_POLYMERASE\_B: 1.  
 KM Transferase: DNA-directed DNA polymerase: DNA replication;  
 KM DNA-binding; Nuclear protein.  
 SO SEQUENCE 1120 AA; 124853 MW; 12FFAAE95EF54FC4 CRC64;  
 OY 1 GRCVQDG 8  
 DB 493 GRCVQDG 500  
 Query Match 74.5%; Score 35; DB 1; Length 1120;  
 Best Local Similarity 75.0%; Pred. NO. 59;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 5  
 PHAC\_SYN6 STANDARD: PRT: 162 AA.  
 AC P13390;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ALLOPHYCOCYANIN ALPHA-B CHAIN.

OS Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=1139;  
 RN [1]  
 RP SEQUENCE.  
 RA Suter F., Fuglistaller P., Lundell D.J., Glazer A.N., Zuber H.;  
 RT "Amino acid sequences of alpha-allophycocyanin B from Synechococcus  
 RT 6301 and Mastigocladus laminosus";  
 RL FEBS Lett. 217:279-282(1987).  
 CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN  
 CC FROM THE PHYCOBILIPROTEIN COMPLEX. ALLOPHYCOCYANIN HAS A MAXIMUM  
 CC ABSORPTION AT APPROXIMATELY 650 NANOMETERS.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.  
 DR PIR: A27398; A27398.  
 DR HSPD: P07119; ICPC.  
 DR InterPro: IPR001659; -  
 DR Pfam: PF00502; Phycobilisome: 1.  
 KM Phycobilisome: Electron transport; Photosynthesis; Bile pigment.  
 FT BINDING 82 82  
 FT PHYCOCYANOBILIN CHROMOPHORE.  
 SQ SEQUENCE 162 AA; 17922 MW; D48E9E083D9BA2C CRC64;  
 OY 2 RCVQDQ 8  
 DB 80 RCVQDQ 86  
 Query Match 72.3%; Score 34; DB 1; Length 162;  
 Best Local Similarity 71.4%; Pred. NO. 15;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 6  
 PACR\_MOUSE STANDARD: PRT: 496 AA.  
 ID P70205;  
 AC P70205;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PITUITARY ADENYLATE CYCLASE ACTIVATING POLYPEPTIDE TYPE I RECEPTOR  
 DE PRECURSOR (PACAP TYPE I RECEPTOR) (PACAP-R-1).  
 GN ADYAPR1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96256640; PubMed=8664310;  
 RA Hashimoto H., Yamamoto K., Hagigara N., Ogawa N., Nishino A.,  
 RA Aino H., Nogi H., Imanishi K., Matsuda T., Baba A.;  
 RT "cDNA cloning of a mouse pituitary adenylate cyclase-activating  
 RT polypeptide receptor";  
 RL Biochim. Biophys. Acta 1281:129-133(1996).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR PACAP-27 AND PACAP-38. THE  
 CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE  
 CC ADENYLATE CYCLASE. MAY REGULATE THE RELEASE OF ADRENOCORTICOTROPIN,  
 CC LUTEINIZING HORMONE, GROWTH HORMONE, PROLACTIN, EPINEPHRINE, AND  
 CC CATECHOLAMINE. MAY PLAY A ROLE IN SPERMATOGENESIS AND SPERM  
 CC MOTILITY. CAUSES SMOOTH MUSCLE RELAXATION AND SECRETION IN THE  
 CC GASTROINTESTINAL TRACT.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D82935; BAA11639.1; -

DR GCRDB; GCR\_1152; -  
 DR MGD; MGI:108449; Adcyap1r1.  
 DR InterPro; IPR000832; -  
 DR InterPro; IPR002285; -  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR PRINTS; PR01156; PACAPRECEPTOR.  
 DR PROSITE; PS00649; G\_PROTEIN\_REC\_F2\_1; 1.  
 DR PROSITE; PS00650; G\_PROTEIN\_REC\_F2\_2; 1.  
 DR PROSITE; PS00651; G\_PROTEIN\_REC\_F2\_3; 1.  
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
 KW Spermatogenesis.  
 FT SIGNAL 1 20  
 FT CHAIN 21 496  
 FT DOMAIN 21 155  
 FT TRANSMEM 156 178  
 FT DOMAIN 179 186  
 FT TRANSMEM 187 205  
 FT DOMAIN 206 227  
 FT TRANSMEM 228 253  
 FT DOMAIN 254 268  
 FT TRANSMEM 269 291  
 FT DOMAIN 292 309  
 FT TRANSMEM 310 332  
 FT DOMAIN 333 378  
 FT TRANSMEM 379 399  
 FT DOMAIN 400 413  
 FT TRANSMEM 414 433  
 FT DOMAIN 434 496  
 FT CARBOHYD 48 48  
 FT CARBOHYD 60 60  
 FT CARBOHYD 117 117  
 FT SEQUENCE 496 AA; 56639 MW; 47B5D51D4209060A CRC64;  
 SO  
 Query Match 72.3%; Score 34; DB 1; Length 496;  
 Best Local Similarity 62.5%; Pred. No. 42;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GRGCVQDG 8  
 Db 115 GRNCTEDG 122  
 RESULT 7  
 ID PACR\_RAT STANDARD; PRT; 523 AA.  
 AC P32215; 063414;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PITUITARY ADENYLATE CYCLASE ACTIVATING POLYPEPTIDE TYPE IA RECEPTOR  
 DE PRECURSOR (PACAP TYPE IA RECEPTOR) (PACAP-R-1).  
 GN ACYAP1R1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Pancreas;  
 RX MEDLINE=93317678; PubMed=8392197;  
 RA Wank S.A., Pilegna J.R.;  
 RT "Molecular cloning and functional expression of the pituitary  
 RT adenylate cyclase-activating polypeptide type I receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6345-6349(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain;  
 RX MEDLINE=93326107; PubMed=7687425;  
 RA Hosoya M., Onda H., Ogi K., Masuda Y., Miyamoto Y., Ohtaki T.,  
 RA Okazaki H., Aritura A., Fujino M.;

FT "Molecular cloning and functional expression of rat cDNAs encoding  
 FT the receptor for pituitary adenylate cyclase activating polypeptide  
 FT (PACAP)." ;  
 FT Biochem. Biophys. Res. Commun. 194:133-143(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Pituitary;  
 RX MEDLINE=93357025; PubMed=8394723;  
 RA Hashimoto H., Ishihara T., Shigemoto R., Mori K., Nagata S.;  
 RT "Molecular cloning and tissue distribution of a receptor for  
 RT pituitary adenylate cyclase-activating polypeptide." ;  
 RL Neuron 11:333-342(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Wistar;  
 RX MEDLINE=93382505; PubMed=8396727;  
 RA Spengler D., Waeber C., Pantaloni C., Holsboer F., Bockaert J.,  
 RA Seeburg P.H., Journot L.;  
 RT "Differential signal transduction by five splice variants of the  
 RT PACAP receptor." ;  
 RL Nature 365:170-175(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Olfactory bulb;  
 RX MEDLINE=93359075; PubMed=8394834;  
 RA Morrow J.A., Lutz E.M., West K.M., Fink G., Harmar A.J.;  
 RT "Molecular cloning and expression of a cDNA encoding a receptor for  
 RT pituitary adenylate cyclase activating polypeptide (PACAP)." ;  
 RL FEBS Lett. 329:99-105(1993).  
 RN [6]  
 RP SEQUENCE OF 115-523 FROM N.A.  
 RC STRAIN-Wistar;  
 RA Svoboda M., Ciccarelli E., Tasteno M., Christophe J.;  
 RL Submitted (MAY-1993) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR PACAP-27 AND PACAP-38. THE  
 CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE  
 CC ADENYLATE CYCLASE. MAY REGULATE THE RELEASE OF ADRENOCORTICOTROPIN,  
 CC LUTEINIZING HORMONE, GROWTH HORMONE, PROLACTIN, EPINEPHRINE, AND  
 CC CATECHOLAMINE. MAY PLAY A ROLE IN SPERMATOGENESIS AND SPERM  
 CC MOTILITY. CAUSES SMOOTH MUSCLE RELAXATION AND SECRETION IN THE  
 CC GASTROINTESTINAL TRACT.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: FIVE ISOFORMS: HIP-HOP1 (SHOWN HERE), HOP1,  
 CC HOP2, HIP AND PACAP-R; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: HYPOTHALAMUS, ANTERIOR PITUITARY, ADRENAL  
 CC MEDULLA, TESTICULAR GERM CELLS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D16465; BAA03932.1; -  
 DR EMBL; L16680; AAA41792.1; -  
 DR EMBL; D14908; BAA03608.1; -  
 DR EMBL; D14909; BAA03609.1; -  
 DR EMBL; Z23272; CAAB0810.1; -  
 DR EMBL; Z23273; CAAB0811.1; -  
 DR EMBL; Z23274; CAAB0812.1; -  
 DR EMBL; Z23275; CAAB0813.1; -  
 DR EMBL; Z23279; CAAB0817.1; -  
 DR EMBL; Z23282; CAAB0821.1; -  
 DR EMBL; Z23282; CAAB0820.1; -  
 DR EMBL; L16506; AAA02990.1; -  
 DR EMBL; Z22735; CAAB0429.1; -  
 DR PIR; S39060; S39060.  
 DR GCRDB; GCR\_0661; -  
 DR GCRDB; GCR\_0662; -  
 DR GCRDB; GCR\_0663; -

```

DR GCRDB: GCR_0664: -
DR GCRDB: GCR_0665: -
DR GCRDB: GCR_0666: -
DR GCRDB: GCR_0667: -
DR GCRDB: GCR_0668: -
DR GCRDB: GCR_0669: -
DR GCRDB: GCR_0670: -
DR GCRDB: GCR_0671: -
DR InterPro: IPR000832: -
DR InterPro: IPR002285: -
DR Pfam: PF00002: 7tm_2; 1.
DR PRINTS: PR00249: GPCRSECRETIN.
DR PRINTS: PR01156: PACAPRECEPTOR.
DR PROSITE: PS00649: G_PROTEIN_RECP_F2_1; 1.
DR PROSITE: PS00650: G_PROTEIN_RECP_F2_2; 1.
DR PROSITE: PS00651: G_PROTEIN_RECP_F2_3; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; signal;
KM Spermatogenesis1: Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 523
FT DOMAIN 20 154
FT TRANSMEM 155 177
FT DOMAIN 178 185
FT TRANSMEM 186 204
FT DOMAIN 205 226
FT TRANSMEM 227 252
FT DOMAIN 253 267
FT TRANSMEM 268 290
FT DOMAIN 291 308
FT TRANSMEM 309 331
FT DOMAIN 332 405
FT TRANSMEM 406 426
FT DOMAIN 427 440
FT TRANSMEM 441 460
FT DOMAIN 461 523
FT CARBOHYD 47 47
FT CARBOHYD 59 59
FT CARBOHYD 116 116
FT VARSPLIC 348 348
FT VARSPLIC 349 376
FT VARSPLIC 349 377
FT VARSPLIC 377 404
FT VARSPLIC 349 404
FT CONFLICT 449 449
FT CONFLICT 510 511
FT CONFLICT 515 516
SO SEQUENCE 523 AA; 59637 MW; 038CA413ED37E44 CRC64;

```

```

Query Match 72.38; Score 34; DB 1; Length 523;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 GRGCVSDG 8
DB 114 GRNCTEDG 121

```

```

RESULT 8
YRS7_CABEL STANDARD; PRT; 796 AA.
ID YRS7_CABEL
AC Q10003;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 90.8 KDA PROTEIN TOSH10.7 IN CHROMOSOME II.
GN TOSH10.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
NCBI_TaxID=6239;

```

```

RN [1]
SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Lightning J., Thomas K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z47811; CA87788.1; JOINED.
DR EMBL: Z47812; CA87788.1; JOINED.
DR EMBL: Z47812; CA87796.1; JOINED.
DR EMBL: Z47811; CA87796.1; JOINED.
DR WormPep: T05H10.7; CE03637.
KM Hypothetical protein.
FT DOMAIN 38 42
FT DOMAIN 67 70
FT DOMAIN 524 527
FT SIGNAL 524 527
SO SEQUENCE 796 AA; 90831 MW; 7BDF8E0A4D2AA9F1 CRC64;

```

```

Query Match 72.38; Score 34; DB 1; Length 796;
Best Local Similarity 75.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 GRGCVSDG 8
DB 198 GRKCVSDG 205

```

```

RESULT 9
DPOU_HCMVA STANDARD; PRT; 1242 AA.
ID DPOU_HCMVA
AC P08546;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN U54 OR HEPF2.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
NCBI_TaxID=10360;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE-87061230; PubMed-3023690;
RA Kouzarides T., Bankier A.T., Satchwell S.C., Weston K.,
RA Barrell B.G.;
RT "Sequence and transcription analysis of the human cytomegalovirus DNA
RT polymerase gene."
RL J. Virol. 61:125-133(1987).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE-87151139; PubMed-3029980;
RA Kouzarides T., Bankier A.T., Satchwell S.C., Weston K.,
RA Tomlinson P., Barrell B.G.;
RT "Large-scale rearrangement of homologous regions in the genomes of
RT HCMV and EBV."
RL Virology 157:397-413(1987).
RN [3]
RN COMPLETE GENOME.
RX MEDLINE-90269039; PubMed-2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohm R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

```

CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -  
 CC N PYROPHOSPHATE + DNA(N).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X17403; CAA35413.1; -  
 CC EMBL: M17209; AAA46008.1; -  
 CC EMBL: M14709; AAA45988.1; -  
 CC PIR: S09817; DJBECL.  
 CC InterPro: IPR002064; -  
 CC Pfam: PF00136; DNA\_POL\_B; 1.  
 CC PRINTS: PR00106; DNAPOLB.  
 CC PROSITE: PS00116; DNA\_POLYMERASE\_B; 1.  
 CC Transferrase: DNA-directed DNA polymerase; DNA replication;  
 CC DNA-binding; Nuclear protein.  
 CC CONFLICT 942 943 943 FT -> VF (IN REF. 3).  
 CC SEQUENCE 1242 AA; 137101 MW; FA59BF842BED80B CRC64;

Query Match 72.3%; Score 34; DB 1; Length 1242;  
 Best Local Similarity 62.5%; Pred. No. 98;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8  
 |||||:  
 DB 347 GRGCTSEG 354

RESULT 10  
 G3P\_CHLPN STANDARD; PRT; 335 AA.  
 AC Q92770; Q92770; (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).  
 GN GAP OR GAP OR CP00624 OR CP0123.  
 OS Chlamydia pneumoniae (Chlamydiophila pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Gilwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RT Nat. Genet. 21:385-389(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Felsen W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39";  
 RT Nucleic Acids Res. 28:1397-1406(2000).  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hatake H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,

RA Shida T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
 CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
 CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
 CC DEHYDROGENASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001647; AAD18763.1; -  
 CC EMBL: AE002173; AAF38006.1; -  
 CC EMBL: AP002547; BAA98831.1; -  
 CC TIGR: CP0123; -  
 CC HSP: P06977; IGAE.  
 CC InterPro: IPR000173; -  
 CC Pfam: PF00044; gpdh; 1.  
 CC PRINTS: PR00078; G3PDHGNASE.  
 CC PROSITE: PS00071; GAPDH; 1.  
 CC Glycolysis; Oxidoreductase; NAD.  
 CC BINDING 151 151 FT  
 FT ACT\_SITE 178 178 FT  
 FT ACTIVATES THIOL GROUP DURING CATALYSIS  
 FT (BY SIMILARITY).  
 FT SEQUENCE 335 AA; 36837 MW; C86D0E3AD3ADF8FE CRC64;

Query Match 70.2%; Score 33; DB 1; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 45;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGCVOD 7  
 |||||:  
 DB 198 GRGCLQN 204

RESULT 11  
 GLGD\_BACST STANDARD; PRT; 343 AA.  
 AC O08327;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE GLYCOGEN BIOSYNTHESIS PROTEIN GLGD.  
 GN GLGD.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TBRE14;  
 RX MEDLINE=97386405; PubMed=9244254;  
 RA Takata H., Takata T., Okada S., Takagi M., Imanaka T.;  
 RT "Characterization of a gene cluster for glycogen biosynthesis and a  
 RT heterotetrameric ADP-glucose pyrophosphorylase from Bacillus  
 RT stearothermophilus";  
 RT J. Bacteriol. 179:4689-4698(1997).  
 RL [2]  
 CC -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF GLYCOGEN.  
 CC -1- SIMILARITY: TO ADP-GLUCOSE PYROPHOSPHORYLASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D87026; BAA19590.1; -

KR Glycogen biosynthesis.

SO SEQUENCE 343 AA; 38965 MW; 20A1370E10A0040 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 343;  
Best Local Similarity 75.0%; Pred. No. 46;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCGVODG 8  
| | | | |

Db 299 GDCGVLDG 306

RESULT 12  
IRF3\_CHICK STANDARD; PRT; 491 AA.

AC 090643;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE INTERFERON REGULATORY FACTOR 3 (IRF-3).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=95334365; PubMed=7541908;  
RA Grant C.E., Vase M.Z., Deeley R.G.;  
RT "cIRF-3, a new member of the interferon regulatory factor (IRF) family that is rapidly and transiently induced by dsRNA.";  
RL Nucleic Acids Res. 23:2137-2146(1995).  
CC -1- FUNCTION: MAY ACTIVATE TRANSCRIPTION BY COMPLEX FORMATION WITH OTHER TRANSCRIPTIONAL FACTORS, POSSIBLY MEMBERS OF THE STAT FAMILY. BINDS SPECIFICALLY TO THE IFN-STIMULATED RESPONSE ELEMENT (ISRE) BUT NOT TO THE IRF-1 BINDING SITE PRD-1.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U20338; AAA86995.1; -

DR HSSP; P15314; 1IF1.  
DR InterPro; IPR001346; -  
DR Pfam; PF00605; IRF.1.  
DR PRINTS; PR00267; INTERFERECT.  
DR PROSITE; PS00601; IRF.1.  
FT Transcription regulation; DNA-binding; Nuclear protein; Activator.  
KW DNA\_BIND 14 112 TRYPTOPHAN PENTAD REPEAT.  
SO SEQUENCE 491 AA; 54441 MW; CAE0C2AA8BE976D9 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 491;  
Best Local Similarity 75.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRCGVODG 8  
| | | | |

Db 235 GCGCGODG 242

RESULT 13  
PIL\_LEPIN STANDARD; PRT; 556 AA.

AC P17627;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SPHINGOMYELINASE C PRECURSOR (EC 3.1.4.12) (SPHINGOMYELIN PHOSPHODIESTERASE) (SMASE).  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEROVAR HARDO / ISOLATE SPONSELEE;  
RX MEDLINE=90307220; PubMed=2163985;  
RA Segers R.P.A.M., van der Drift A., de Nijls A., Corcione P., van der Zeijst B.A.M., Gaastra W.;  
RT "Molecular analysis of a sphingomyelinase C gene from Leptospira interrogans serovar hardjo.";  
RL Infect. Immun. 58:2177-2185(1990).  
CC -1- CATALYTIC ACTIVITY: SPHINGOMYELIN + H(2)O -> N-ACYLSPHINGOSINE + CHOLINE PHOSPHATE.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- SIMILARITY: TO BACILLUS SPHINGOMYELINASES.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X52176; CAA36424.1; -

DR PIR; S22634; S22634.  
KW Hydrolyase; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 556 SPHINGOMYELINASE C.  
SO SEQUENCE 556 AA; 63268 MW; 6D0EACDB9A5CD6D CRC64;

Query Match 70.2%; Score 33; DB 1; Length 556;  
Best Local Similarity 83.3%; Pred. No. 71;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 GCVODG 8  
| | | | |

Db 489 GCLODG 494

RESULT 14  
YC92\_CAEEL STANDARD; PRT; 621 AA.

AC P55115;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL ZINC METALLOPROTEINASE T04G9.2 (EC 3.4.24.-).  
GN T04G9.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Pelodermidae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Mink P.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC

```

CC METALLOPROTEASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U41274; AA82461.1; -.
CC DR HSSP: P07584; 1IAD.
CC DR WormPep: T04G9-2; CE04883.
CC DR InterPro: IPR000130; -.
CC DR InterPro: IPR001506; -.
CC DR Pfam: PF01400; Astacin; 1.
CC DR PRINTS: PR00480; ASTACIN.
CC DR PROSITE: PS00142; ZINC_PROTEASE; 1.
CC DR Hypothetical protein: Hydrolase; Metalloprotease; zinc.
CC KW METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 256 256 BY SIMILARITY.
CC FT METAL 259 259 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
CC SQ SEQUENCE 621 AA; 68584 MW; A98D3721C8ADE489 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 621;
Best Local Similarity 62.5%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GRGCVODG 8
DB 244 GSGCIQKG 251

RESULT 15
TKT1_KIULA STANDARD; PRT; 679 AA.
AC 012630;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSKETOLASE (EC 2.2.1.1) (Tk).
GN TKL1.
OS Kluveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CBS 2359/152;
RX MEDLINE=97153463; PubMed=9000376;
RA Jacoby J.J., Heinisch J.J.;
RT "Analysis of a transketolase gene from Kluveromyces lactis reveals
RT that the yeast enzymes are more related to transketolases of
RT prokaryotic origins than to those of higher eukaryotes.";
RL Curr. Genet. 31:15-21(1997).
CC -1- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC 3-PHOSPHATE -> D-RIBOSE 5-PHOSPHATE + D-XYULOSE 5-PHOSPHATE.
CC -1- COFACTOR: THIAMINE PYROPHOSPHATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U65983; AAB05935.1; -.
CC DR HSSP: P23234; IAY0.

```

```

DR InterPro: IPR000360; -.
DR Pfam: PF00456; transketolase; 1.
DR PROSITE: PS00801; TRANSKETOLASE_1; 1.
DR PROSITE: PS00802; TRANSKETOLASE_2; 1.
KW Transferrase; Thiamine pyrophosphate.
SQ SEQUENCE 679 AA; 73703 MW; BC5F3CF61A78CD4A CRC64;

Query Match 70.2%; Score 33; DB 1; Length 679;
Best Local Similarity 62.5%; Pred. No. 85;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVODG 8
DB 156 GDCIQEG 163

Search completed: June 13, 2001, 14:21:44
Job time: 804 sec

```



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:29 ; Search time 225.85 Seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-10

Perfect score: 47

Sequence: 1 GRCVODG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_podent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	78.7	250	10	09LGT0
2	37	78.7	323	2	09LRF5
3	37	78.7	399	2	09R3V6
4	37	78.7	399	2	09L7G6
5	37	78.7	653	5	09VZK2
6	37	78.7	664	5	024140
7	37	78.7	1218	5	09VHR6
8	36	76.6	206	13	090802
9	36	76.6	379	2	09K162
10	36	76.6	814	10	065715
11	36	76.6	1081	5	09W3R9
12	36	76.6	1267	5	09U0U5
13	35	74.5	321	5	076213
14	35	74.5	346	4	09S274
15	35	74.5	346	4	09U074
16	35	74.5	352	11	055162
17	35	74.5	449	5	026978
18	35	74.5	983	3	094165
19	34	72.3	255	5	09NBC9

20	34	72.3	296	2	050407	050407 mycobacteri
21	34	72.3	329	2	092HK1	092HK1 rhodococcus
22	34	72.3	757	5	09VNF8	09VNF8 drosophila
23	34	72.3	769	5	09NFN9	09NFN9 drosophila
24	34	72.3	1240	14	09WHP1	09WHP1 human cytom
25	34	72.3	1242	14	09WH06	09WH06 human cytom
26	34	72.3	1242	14	09WH05	09WH05 human cytom
27	34	72.3	1242	14	09WH04	09WH04 human cytom
28	34	72.3	1242	14	09WH03	09WH03 human cytom
29	34	72.3	1242	14	09WH02	09WH02 human cytom
30	34	72.3	1242	14	09WH01	09WH01 human cytom
31	34	72.3	1242	14	09WHP9	09WHP9 human cytom
32	34	72.3	1242	14	09WHP8	09WHP8 human cytom
33	34	72.3	1242	14	09WHP7	09WHP7 human cytom
34	34	72.3	1242	14	09WHP6	09WHP6 human cytom
35	34	72.3	1242	14	09WHP5	09WHP5 human cytom
36	34	72.3	1242	14	09WHP4	09WHP4 human cytom
37	34	72.3	1242	14	09WHP3	09WHP3 human cytom
38	34	72.3	1242	14	09WHP2	09WHP2 human cytom
39	34	72.3	1242	14	09WHP0	09WHP0 human cytom
40	34	72.3	1242	14	09WHN9	09WHN9 human cytom
41	34	72.3	1242	14	09WHN8	09WHN8 human cytom
42	34	72.3	1242	14	09WHN7	09WHN7 human cytom
43	34	72.3	1242	14	09WHN6	09WHN6 human cytom
44	34	72.3	1242	14	09WHN5	09WHN5 human cytom
45	34	72.3	1242	14	09WHN0	09WHN0 human cytom

## ALIGNMENTS

RESULT 1	09LGT0	PRELIMINARY:	PRT:	250 AA.
ID	09LGT0			
AC	09LGT0			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	P0489A01.11 PROTEIN.			
GN	P0489A01.11.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
CC	Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.			
OX	NCBI_TaxID=4530;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. NIPPONBARE;			
RA	Sasaki T., Matsumoto T., Yamamoto K.;			
RT	"Oryza sativa nipponbare(GN3) genomic DNA, chromosome 1, PAC clone:P0489A01."			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AP002484; BAA99520.1; -			
SQ	SEQUENCE 250 AA; 28494 MW; 27574C02BA14BCF3 CRC64;			
Query Match	78.7%;	Score 37;	DB 10;	Length 250;
Best Local Similarity	75.0%;	Pred. No. 12;		
Matches	6;	Conservative	1;	Mismatches
			1;	Indels
				Gaps
				0;
Oy	1 GRCVODG 8			
DB	17 GAGCYKDG 24			
RESULT 2	09LRF5	PRELIMINARY:	PRT:	323 AA.
ID	09LRF5			
AC	09LRF5			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	ZOT. (FRAGMENT).			
GN	ZOT.			

```

OS   Vibrio mimicus.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=674;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   MEDLINE=20143766; PubMed=10678967;
RA   Boyd E.F., Meyer K.E., Shi L., Waldor M.K.;
RT   "Infectious CTXphi and the vibrio pathogenicity island prophage in
RT   Vibrio mimicus: evidence for recent horizontal transfer between V.
RL   mimicus and V. cholerae.";
DR   EMBL, AF207857; AAF40142.1; -.
FT   NON_TER
FT   NON_TER
FT   NON_TER
SQ   SEQUENCE 323 AA; 36306 MM; 01C12DAEB9B873C3B CRC64;

Query Match      78.7%; Score 37; DB 2; Length 323;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY   1 GRGCVODG 8
Db   240 GRGCVODG 247

RESULT 3
O9R3V6 PRELIMINARY; PRT; 399 AA.
AC O9R3V6;
ID O9R3V6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE ZONULAR OCCLUDENS TOXIN (ZONA OCCLUDENS TOXIN).
GN zot OR VC1458.
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNH002;
RA Shin H.J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RT   cassette from vibrio cholerae KNH002 isolated in Korea.";
RL Misalimurhag Holji 35:203-210(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-0139-TOR OGAWA;
RA Zhi-Yong H., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and Expression of zot Gene from Vibrio cholerae.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL Nature 406:477-483(2000).
DR   EMBL, AF175708; RAD51358.1; -.
DR   EMBL, AF123049; AAD26854.1; -.
DR   EMBL, AE004224; AAF94615.1; -.
DR   TIGR, VC1458; -.
SQ   SEQUENCE 399 AA; 44903 MM; 3C7424B758176774 CRC64;

```

```

Query Match      78.7%; Score 37; DB 2; Length 399;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY   1 GRGCVODG 8
Db   291 GRGCVODG 298

RESULT 4
O9L706 PRELIMINARY; PRT; 399 AA.
AC O9L706;
ID O9L706;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE zot.
GN zot.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-86015;
RA Kan B., Liu Y.Q., Qi G.M., Gao S.Y.;
RT "Vibrio cholerae nct-CTXphi whole genome, include rstr(Rstr),
RT   rsta(Rsta), rscB(RstB), cep(Cep), orf1(Orf1), ace(Ace) and zot(zot)
RT   genes.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR   EMBL, AF220606; AAF29547.1; -.
SQ   SEQUENCE 399 AA; 44990 MM; CFA3DBCC9CE23BE1 CRC64;

Query Match      78.7%; Score 37; DB 2; Length 399;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY   1 GRGCVODG 8
Db   291 GRGCVODG 298

RESULT 5
O9VZK2 PRELIMINARY; PRT; 653 AA.
AC O9VZK2;
ID O9VZK2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SCRT PROTEIN (SCRATCH).
GN SCRT.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Piankoff C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fioder C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirsky R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN
RP SEQUENCE OF 1-567 FROM N.A.
RA Rudin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcalini T.T., Baxter E., Blazej R.G., Butenhoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
RA Park S., Sequerra A., Sethi H., Snir E., Svirsky R.R., Weinburg T.,
RA Ceinlier S.E.;
RA "Full length Drosophila melanogaster cDNA sequence."
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003479; AAF47819.1; -
DR EMBL: AF145627; AAD38602.1; -
DR HSSP: P25490; IUBD.
DR FLYBASE: FBgn0004880; scrt.
DR INTERPRO: IPR000822; -
DR PFAM: PF00096; zf-C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 4.
KW Zinc-finger: Metal-binding; DNA-binding.
SQ SEQUENCE 653 AA: 70774 MW: 15C27EE93F3FF45D CRC64:

Query Match      78.7%; Score 37; DB 5; Length 653;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8
   1 11:111
DB 408 GDGCLQDG 415

RESULT 6
ID 024140 PRELIMINARY; PRT; 664 AA.
AC 024140;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE NEURON SPECIFIC ZINC FINGER TRANSCRIPTION FACTOR.
GN SCRATCH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96010286; PubMed=7557390;

```

```

RA Roark M., Sturtevant M.A., Emery J., Vaessin H., Grell E., Bier E.;
RT "scratch, a pan-neural gene encoding a zinc finger protein related to
RT snail, promotes neuronal development."
RL Genes Dev. 9:2384-2398(1995).
DR EMBL: U36477; AAA91035.1; -
DR HSSP: P25490; IUBD.
DR FLYBASE: FBgn0004880; scrt.
DR INTERPRO: IPR000822; -
DR PFAM: PF00096; zf-C2H2; 5.
DR PROSITE: PR00048; ZINC_FINGER.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 4.
KW Zinc-finger: Metal-binding; DNA-binding.
SQ SEQUENCE 664 AA: 72569 MW: 687F2DAABFE8968F CRC64:

Query Match      78.7%; Score 37; DB 5; Length 664;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGCVQDG 8
   1 11:111
DB 408 GDGCLQDG 415

RESULT 7
ID 09V8R6 PRELIMINARY; PRT; 1218 AA.
AC 09V8R6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE CG7097 PROTEIN.
GN CG7097.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceinlier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Brattin J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fioder C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirsky R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Svirsky R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003796; AAF57595.1; -  
 DR HSSP: P24941; IAO1.  
 DR FLYBASE: FBgn0034421; CG7097.  
 DR INTERPRO: IPR00719; -  
 DR INTERPRO: IPR01180; -  
 DR PFAM: PF00069; Kinase; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 SO SEQUENCE 1218 AA; 122395 MW; 78A/FAFF880CEDC8 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 1218;  
 Best Local Similarity 75.0%; Pred. No. 56;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCVQDG 8  
 DB 1024 GHGVQNG 1031

RESULT 8  
 ID 090802 PRELIMINARY; PRT; 206 AA.  
 AC 090802;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE COLLAGEN TYPE 2 PRECURSOR (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92078225; PubMed=1744138;  
 RA Nab H.D., Upholt W.B.,  
 RT "Type II collagen mRNA containing an alternatively spliced exon  
 RT predominates in the chick limb prior to chondrogenesis,"  
 RL J. Biol. Chem. 266:23446-23452(1991).  
 DR EMBL: M74435; AAA48714.1; -  
 DR INTERPRO: IPR000087; -  
 DR INTERPRO: IPR001007; -  
 DR PRAM: PR00093; VWC; 1.  
 DR PRAM: PF01381; Collagen; 1.  
 DR PROSITE: PS01208; VWC; 1.  
 DR Signal.  
 FT SIGNAL 185 >206 POTENTIAL.  
 FT CHAIN 206 COLLAGEN TYPE 2.  
 FT NON TER 206  
 SO SEQUENCE 206 AA; 20793 MW; E3EB84ACF0C31459 CRC64;

Query Match 76.6%; Score 36; DB 13; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GCVQDG 8  
 DB 36 GCVQDG 41

RESULT 9  
 O9KY62 PRELIMINARY; PRT; 379 AA.  
 ID O9KY62

AC O9KY62;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 39.4 KDA PROTEIN.  
 GN DP63A.13C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetaceae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome,"  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL356832; CAB92665.1; -  
 SO SEQUENCE 379 AA; 39436 MW; CD9855130DE03A76 CRC64;

Query Match 76.6%; Score 36; DB 2; Length 379;  
 Best Local Similarity 62.5%; Pred. No. 28;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRCVQDG 8  
 DB 46 GRCVQDG 53

RESULT 10  
 ID 065715 PRELIMINARY; PRT; 814 AA.  
 AC 065715;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 91.9 KDA PROTEIN.  
 GN T5K18.220 OR AT4G19440.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Van Der Schueren J., Chuang Y.-J., Voet M., Robben J.,  
 RA Volckaert G., Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;  
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Van Der Schueren J., Chuang Y.-J., Aert R., Defoor E., Robben J.,  
 RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL022580; CAA18631.1; -  
 DR EMBL: AL161550; CAB78946.1; -  
 DR INTERPRO: IPR002885; -

DR PFAM: PF01535; PPR: 17.  
 KM Hypothetical protein.  
 SO SEQUENCE 814 AA; 91943 MW; F5BFE318F5A05B36 CRC64;

Query Match 76.6%; Score 36; DB 10; Length 814;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRCVOD 7  
 |||||  
 DB 522 GRCVMD 528

RESULT 11  
 09W3R9 PRELIMINARY; PRT; 1081 AA.  
 ID 09W3R9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE CG1677 PROTEIN.  
 GN CG1677.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RC STRAIN=BERKELEY;  
 RA Adams M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-U., Andrews-Plannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher S., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklav G., Mishina N.V., Moberly C., Morris J., Mosnelli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacble J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spelling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003440; AAF46248.1;  
 DR FLYBASE: FBgn0029941; CG1677.  
 SO SEQUENCE 1081 AA; 118112 MW; FA50144690473098 CRC64;

Query Match 76.6%; Score 36; DB 5; Length 1081;  
 Best Local Similarity 62.5%; Pred. No. 78;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRCVOD 8  
 |||||  
 DB 432 GNCLEDDG 439

RESULT 12  
 09U0U5 PRELIMINARY; PRT; 1267 AA.  
 ID 09U0U5;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE HYPOTHETICAL 132.4 KDA PROTEIN.  
 GN L8980.02.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 NC NCB1\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIEDLIN;  
 RA Wedler H., Hilbert H., Duesterhoeft A., Ivens A.C., Murphy L.,  
 RA Quail M., Rajandream M.A., Barrell B.G.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIEDLIN;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome.";  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL: AL135958; CAB64588.1;  
 KW Hypothetical protein.  
 SO SEQUENCE 1267 AA; 132368 MW; F5EE59A830E4946B CRC64;

Query Match 76.6%; Score 36; DB 5; Length 1267;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCVOD 8  
 |||||  
 DB 466 GCVODG 471

RESULT 13  
 076213 PRELIMINARY; PRT; 321 AA.  
 ID 076213;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE REVERSE TRANSCRIPTASE (FRAGMENT).  
 GN POL.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Neodermata; Trematoda; Digenea;  
 OC Strigoida; Schistosomatidae; Schistosomatidae; Schistosoma.  
 NC NCB1\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Drew A.C.;  
 RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AF023673; AAC24983.1;  
 KW RNA-directed DNA polymerase.  
 FT NON\_TER 1  
 FT NON\_TER 321  
 SO SEQUENCE 321 AA; 35998 MW; E267170D26B6388 CRC64;

Query Match 74.5%; Score 35; DB 5; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: June 13, 2001, 14:20:30  
 Job time: 730 sec

OY 2 RGCYOD 7  
 |||||  
 Db 186 RGCYOD 191

## RESULT 14

O95274 PRELIMINARY; PRT; 346 AA.  
 AC O95274;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE GPI-ANCHORED METASTASIS-ASSOCIATED PROTEIN HOMOLOG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith B., Southgate J.;  
 RT "Differentially expressed gene in urothelial cell:matrix  
 interactions.";  
 RT Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF082889; AAD13751.1;  
 SQ SEQUENCE 346 AA; 35924 MW; CDFP9948FF493130 CRC64;

Query Match 74.5%; Score 35; DB 4; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGCYOD 7  
 |||||  
 Db 191 RGCYOD 196

## RESULT 15

O9UJ74 PRELIMINARY; PRT; 346 AA.  
 AC O9UJ74;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE HYPOTHETICAL 36.0 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hassel B., Reimann M., Wuerfel J., Claas A., Savelyeva L., Schwab M.,  
 RA Roessel M., Matzku S., Zoeller M.;  
 RT "Cloning of the human homologue of the metastasis-associated rat  
 C4.4A.";  
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AJ223603; CA11469.2;  
 KW Hypothetical protein.  
 SQ SEQUENCE 346 AA; 35970 MW; 97FF9B4A554934FF CRC64;

Query Match 74.5%; Score 35; DB 4; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RGCYOD 7  
 |||||  
 Db 191 RGCYOD 196



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:34 ; Search time 118.55 Seconds

(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-10

Perfect score: 47

Sequence: 1 GRGCVQDG 8

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	80.9	176	2 US-08-835-231-18	Sequence 18, Appl
2	37	78.7	400	2 US-08-624-601-8	Sequence 8, Appl
3	34	72.3	448	2 US-08-811-897A-18	Sequence 18, Appl
4	34	72.3	448	2 US-08-855-213-18	Sequence 18, Appl
5	34	72.3	467	2 US-08-811-897A-19	Sequence 19, Appl
6	34	72.3	457	2 US-08-855-213-19	Sequence 19, Appl
7	34	72.3	476	2 US-08-811-897A-20	Sequence 20, Appl
8	34	72.3	476	2 US-08-855-213-20	Sequence 20, Appl
9	34	72.3	495	2 US-08-811-897A-21	Sequence 21, Appl
10	34	72.3	495	2 US-08-855-213-21	Sequence 21, Appl
11	34	72.3	1242	2 US-08-680-326-33	Sequence 33, Appl
12	31	66.0	295	4 US-08-358-368-2	Sequence 2, Appl
13	31	66.0	376	2 US-08-558-269-10	Sequence 10, Appl
14	31	66.0	449	1 US-08-142-439A-5	Sequence 5, Appl
15	31	66.0	449	2 US-08-469-477-5	Sequence 5, Appl
16	31	66.0	579	1 US-08-295-411-4	Sequence 4, Appl
17	31	66.0	579	2 US-08-955-471-4	Sequence 4, Appl
18	31	66.0	579	3 US-09-117-708-14	Sequence 14, Appl
19	31	66.0	579	5 PCT-US92-10242-4	Sequence 4, Appl
20	31	66.0	615	1 US-07-998-972A-3	Sequence 3, Appl
21	31	66.0	615	1 US-08-463-953-3	Sequence 3, Appl
22	31	66.0	615	1 US-08-462-261-3	Sequence 3, Appl
23	31	66.0	615	5 PCT-US92-11357-3	Sequence 3, Appl
24	31	66.0	622	3 US-08-952-967-8	Sequence 8, Appl
25	31	66.0	816	1 US-08-038-760-3	Sequence 3, Appl
26	31	66.0	816	2 US-08-470-091-3	Sequence 3, Appl
27	31	66.0	928	1 US-08-204-329-1	Sequence 1, Appl

28	31	66.0	928	2 US-08-959-638-8	Sequence 8, Appl
29	31	66.0	928	2 US-08-482-627-5	Sequence 5, Appl
30	31	66.0	928	3 US-08-801-092-4	Sequence 4, Appl
31	31	66.0	928	4 US-08-328-673A-8	Sequence 8, Appl
32	31	66.0	928	5 PCT-US94-10357-2	Sequence 2, Appl
33	31	66.0	928	5 PCT-US94-10357-3	Sequence 3, Appl
34	31	66.0	2052	4 US-09-045-201A-2	Sequence 2, Appl
35	30	63.8	192	1 US-08-086-428B-86	Sequence 86, Appl
36	30	63.8	192	1 US-08-086-428B-87	Sequence 87, Appl
37	30	63.8	192	1 US-08-086-428B-88	Sequence 88, Appl
38	30	63.8	192	1 US-08-086-428B-89	Sequence 89, Appl
39	30	63.8	192	1 US-08-086-428B-90	Sequence 90, Appl
40	30	63.8	192	2 US-08-468-570-86	Sequence 86, Appl
41	30	63.8	192	2 US-08-468-570-87	Sequence 87, Appl
42	30	63.8	192	2 US-08-468-570-88	Sequence 88, Appl
43	30	63.8	192	2 US-08-468-570-89	Sequence 89, Appl
44	30	63.8	192	2 US-08-468-570-90	Sequence 90, Appl
45	30	63.8	192	2 US-08-290-665A-86	Sequence 86, Appl

## ALIGNMENTS

RESULT 1  
US-08-835-231-18  
Sequence 18, Application US/08835231  
Patent No. 5861284  
GENERAL INFORMATION:  
APPLICANT: NISHIMURA, Osamu  
APPLICANT: KURIYAMA, Masato  
APPLICANT: KOYAMA, No. 5861284YUK1  
APPLICANT: FUKUDA, Tsunehiko  
TITLE OF INVENTION: METHOD FOR PRODUCING A BIOLOGICALLY  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 WATER STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/835,231  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,709  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: 07/838,857  
FILING DATE: 18-FEB-1992  
APPLICATION NUMBER: JP 024841  
FILING DATE: 19-FEB-1991  
APPLICATION NUMBER: JP 0271438  
FILING DATE: 18-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: DAVID, RESNICK S  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 41614-FWC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX: 200291 STRE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 176 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-835-231-18

Query Match 80.9%; Score 38; DB 2; Length 176;  
Best Local Similarity 75.0%; Pred. No. 15;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVODG 8  
1111:11  
DB 30 GRGCPEDG 37

RESULT 2  
US-08-624-601-8  
Sequence 8, Application US/08624601  
Patent No. 5882653  
GENERAL INFORMATION:  
APPLICANT: Kaper Dr., James B.  
TITLE OF INVENTION: Vibrio cholerae O1 (CVD11) and non-O1  
TITLE OF INVENTION: (CVD112 and CVD112RM) serogroup vaccine strains, methods  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spencer & Frank  
STREET: 1100 New York Ave. N.W. Suite 300 East  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/624,601  
FILING DATE: 08-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Schneller Dr., John W.  
REGISTRATION NUMBER: 26,031  
REFERENCE/DOCKET NUMBER: BACZ0019P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)414-4000  
TELEFAX: (202)414-4040  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 amino acids  
TYPE: amino acid  
STRANDEDNESS: US  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Vibrio cholerae  
STRAIN: El Tor 7946  
IMMEDIATE SOURCE:  
CLONE: zot  
US-08-624-601-8

Query Match 78.7%; Score 37; DB 2; Length 400;  
Best Local Similarity 87.5%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGCVODG 8  
11111111  
DB 291 GRGCVODG 298

RESULT 3  
US-08-811-897A-18  
Sequence 18, Application US/08811897A  
Patent No. 5858787  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSoya, Masaki  
APPLICANT: OGI, Kazuhiro  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 585878710  
TITLE OF INVENTION: PACAP RECEPTER PROTEIN, METHOD FOR PREPARING  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,897A  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: February 25, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-9400  
TELEFAX: (617)523-6440  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 448 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-897A-18

Query Match 72.3%; Score 34; DB 2; Length 448;  
Best Local Similarity 62.5%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVODG 8  
11111111  
DB 95 GRGCTEDG 102

RESULT 4  
US-08-855-213-18  
Sequence 18, Application US/08855213  
Patent No. 5892004  
GENERAL INFORMATION:

APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOTA, Masaki  
APPLICANT: OGI, Kazuhiko  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 589200410  
TITLE OF INVENTION: PACAP RECEPTOR PROTEIN, METHOD FOR  
NUMBER OF SEQUENCES: 55  
TITLE OF INVENTION: PREPARING SAID PROTEIN, AND USE THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,213  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: 25-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 448 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-855-213-18

Query Match 72.3%; Score 34; DB 2; Length 448;  
Best Local Similarity 62.5%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GRGCVODG 8  
DB 95 GRNCTEDG 102  
RESULT 5  
US-08-811-897A-19  
Sequence 19, Application US/08811897A  
Patent No. 5858787  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOTA, Masaki  
APPLICANT: OGI, Kazuhiko  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo

APPLICANT: SHIMAMOTO, No. 585878710  
TITLE OF INVENTION: PACAP RECEPTOR PROTEIN, METHOD FOR PREPARING  
NUMBER OF SEQUENCES: 56  
TITLE OF INVENTION: SAID PROTEIN, AND USE THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,897A  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: February 25, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 467 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-897A-19

Query Match 72.3%; Score 34; DB 2; Length 467;  
Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GRGCVODG 8  
DB 114 GRNCTEDG 121  
RESULT 6  
US-08-855-213-19  
Sequence 19, Application US/08855213  
Patent No. 5892004  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOTA, Masaki  
APPLICANT: OGI, Kazuhiko  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 589200410  
TITLE OF INVENTION: PACAP RECEPTOR PROTEIN, METHOD FOR  
NUMBER OF SEQUENCES: 55  
TITLE OF INVENTION: PREPARING SAID PROTEIN, AND USE THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts

COUNTRY: .US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,213  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: 25-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 467 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-855-213-19

Query Match 72.3%; Score 34; DB 2; Length 467;  
Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVODG 8  
DB 114 GRNCTEDG 121

RESULT 7  
US-08-811-897A-20  
Sequence 20, Application US/08811897A  
Patent No. 5858787  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOYA, Masaki  
APPLICANT: OGI, Kazuhiko  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 585878710  
TITLE OF INVENTION: PACAP RECEPTOR PROTEIN, METHOD FOR PREPARING  
TITLE OF INVENTION: SAID PROTEIN, AND USE THEREOF  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,897A

FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: February 25, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-897A-20

Query Match 72.3%; Score 34; DB 2; Length 476;  
Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVODG 8  
DB 95 GRNCTEDG 102

RESULT 8  
US-08-855-213-20  
Sequence 20, Application US/08855213  
Patent No. 5892004  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOYA, Masaki  
APPLICANT: OGI, Kazuhiko  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 589200410  
TITLE OF INVENTION: PACAP RECEPTOR PROTEIN, METHOD FOR  
TITLE OF INVENTION: PREPARING SAID PROTEIN, AND USE THEREOF  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,213  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: 25-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-855-213-20

Query Match 72.3% Score 34; DB 2; Length 476;  
Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
111111  
DB 95 GRNCTEDG 102

RESULT 9  
US-08-811-897A-21  
Sequence 21, Application US/08811897A  
Patent No. 5858787  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOYA, Masaki  
APPLICANT: OGI, Kazuhiro  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 585878710  
TITLE OF INVENTION: PACAP RECEPTER PROTEIN, METHOD FOR PREPARING  
TITLE OF INVENTION: SAID PROTEIN, AND USE THEREOF  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,897A  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: February 25, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-897A-21

Query Match 72.3% Score 34; DB 2; Length 495;  
Best Local Similarity 62.5%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
111111  
DB 114 GRNCTEDG 121

RESULT 10  
US-08-855-213-21  
Sequence 21, Application US/08855213  
Patent No. 3892004  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHTAKI, Tetsuya  
APPLICANT: MASUDA, Yasushi  
APPLICANT: KITADA, Chieko  
APPLICANT: ISHIBASHI, Yoshihiro  
APPLICANT: HOSOYA, Masaki  
APPLICANT: OGI, Kazuhiro  
APPLICANT: MIYAMOTO, Yasunori  
APPLICANT: HABATA, Yugo  
APPLICANT: SHIMAMOTO, No. 589200410  
TITLE OF INVENTION: PACAP RECEPTER PROTEIN, METHOD FOR  
TITLE OF INVENTION: PREPARING SAID PROTEIN, AND USE THEREOF  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,213  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/202,986  
FILING DATE: 25-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 44168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-855-213-21

Query Match 72.3% Score 34; DB 2; Length 495;  
Best Local Similarity 62.5%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
111111  
DB 114 GRNCTEDG 121

RESULT 11  
US-08-680-326-33  
Sequence 33, Application US/08680326  
Patent No. 5925733  
GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARINIX  
APPLICANT: STRAND, KURT  
APPLICANT: TODARO, GEORGE J.  
TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES  
TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL  
NUMBER OF SEQUENCES: 152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOENSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,326  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Schilff, J. Michael  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1242 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-680-326-33

Query Match 72.3%; Score 34; DB 2; Length 1242;  
Best Local Similarity 62.5%; Pred. No. 4.5e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
|||||  
Db 347 GRGCTSEG 354

RESULT 12  
US-08-338-368-2  
Sequence 2, Application US/08338368  
Patent No. 6110721  
GENERAL INFORMATION:  
APPLICANT: GIBBS, CRAIG S.  
APPLICANT: LEUNG, LAWRENCE L.K.  
APPLICANT: TSIANG, MANUEL  
TITLE OF INVENTION: NOVEL POLYPEPTIDES AND COAGULATION  
TITLE OF INVENTION: THERAPY  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GILEAD SCIENCES, INC.  
STREET: 353 LAKESIDE DRIVE  
CITY: FOSTER CITY  
STATE: CALIFORNIA

COUNTRY: USA  
ZIP: 94404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,368  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,038  
FILING DATE: 10-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: HENSLEY, MAX D.  
REGISTRATION NUMBER: 27,043  
REFERENCE/DOCKET NUMBER: 190.2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-574-3000  
TELEFAX: 415-573-4899  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 295 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-338-368-2

Query Match 66.0%; Score 31; DB 4; Length 295;  
Best Local Similarity 62.5%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GRGCVQDG 8  
|||||  
Db 264 GRGCDRDG 271

RESULT 13  
US-08-558-269-10  
Sequence 10, Application US/08558269  
Patent No. 5961973  
GENERAL INFORMATION:  
APPLICANT: Crea, Roberto  
TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/558,269  
FILING DATE: 13-NOV-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/847,800  
FILING DATE: 06-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: CRI-001CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 376 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-558-269-10

Query Match 66.0%; Score 31; DB 2; Length 376;  
Best Local Similarity 62.5%; Pred. No. 4.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGCYVDG 8  
DB 345 RGCYVDG 352

RESULT 14  
US-08-142-439A-5  
Sequence 5, Application US/08142439A  
Patent No. 5670360  
GENERAL INFORMATION:  
APPLICANT: Thorens, Bernard  
TITLE OF INVENTION: Receptor for the Glucagon-Like-Peptide-1  
TITLE OF INVENTION: (GLP-1)  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 56703600 No. 5670360disk of No. 5670360th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/142.439A  
FILING DATE: 24-NOV-93  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 398/92  
FILING DATE: 25-MAR-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/00697  
FILING DATE: 23-MAR-93  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 3756.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 449 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
STRAIN: Sprague-Dawley  
US-08-142-439A-5

Query Match 66.0%; Score 31; DB 1; Length 449;  
Best Local Similarity 71.4%; Pred. No. 5.1e+02;

Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 RGCYVDG 8  
DB 105 RNCYVDG 111

RESULT 15  
US-08-869-477-5  
Sequence 5, Application US/08869477  
Patent No. 5846747  
GENERAL INFORMATION:  
APPLICANT: Thorens, Bernard  
TITLE OF INVENTION: Receptor for the Glucagon-Like-Peptide-1  
TITLE OF INVENTION: (GLP-1)  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 58467470 No. 5846747disk of No. 5846747th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/869.477  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/142.439  
FILING DATE: 24-NOV-93  
APPLICATION NUMBER: DK 398/92  
FILING DATE: 25-MAR-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/00697  
FILING DATE: 23-MAR-93  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 3756.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 449 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
STRAIN: Sprague-Dawley  
US-08-869-477-5

Query Match 66.0%; Score 31; DB 2; Length 449;  
Best Local Similarity 71.4%; Pred. No. 5.1e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RGCYVDG 8  
DB 105 RNCYVDG 111

Search completed: June 13, 2001, 14:16:34

Job time: 495 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:14:35 ; Search time 229.28 Seconds

(without alignments)  
1.995 Million cell updates/sec

Title: PCT-US01-05825A-11

Perfect score: 43  
Sequence: 1 GRGLVQPG 8Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Database :

A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	8	21	Y79115
2	43	100.0	246	14	R40924
3	38	88.4	8	21	Y79113
4	38	88.4	116	16	R66349
5	38	88.4	240	22	B45991
6	38	88.4	240	22	B45992
7	38	88.4	240	22	B45993
8	38	88.4	240	22	B45994
9	38	88.4	240	22	B45995
10	38	88.4	240	22	B45996
11	38	88.4	240	22	B45997

## ALIGNMENTS

12	38	88.4	240	22	B45998	Human MUC-1 scFv C
13	38	88.4	240	22	B45999	Human MUC-1 scFv C
14	38	88.4	240	22	B46000	Human MUC-1 scFv C
15	38	88.4	240	22	B46001	Human MUC-1 scFv C
16	38	88.4	240	22	B46003	Human MUC-1 scFv C
17	38	88.4	240	22	B46004	Human MUC-1 scFv C
18	38	88.4	240	22	B46005	Human MUC-1 scFv C
19	38	88.4	240	22	B46006	Human MUC-1 scFv C
20	38	88.4	240	22	B46007	Human MUC-1 scFv C
21	38	88.4	240	22	B46008	Human MUC-1 scFv C
22	38	88.4	240	22	B46009	Human MUC-1 scFv C
23	38	88.4	240	22	B46010	Human MUC-1 scFv C
24	38	88.4	240	22	B46011	Human MUC-1 scFv C
25	38	88.4	240	22	B46012	Human MUC-1 scFv C
26	38	88.4	240	22	B46013	Human MUC-1 scFv C
27	38	88.4	240	22	B46014	Human MUC-1 scFv C
28	38	88.4	240	22	B46015	Human MUC-1 scFv C
29	38	88.4	240	22	B46016	Human MUC-1 scFv C
30	38	88.4	240	22	B46017	Human MUC-1 scFv C
31	38	88.4	240	22	B46019	Human MUC-1 scFv C
32	38	88.4	240	22	B46020	Human MUC-1 scFv C
33	38	88.4	240	22	B46021	Human MUC-1 scFv C
34	37	86.0	118	18	W19878	CEA-specific antib
35	37	86.0	154	21	B32388	Human secreted pro
36	37	86.0	225	20	Y41165	Human Vhh polypept
37	37	86.0	928	21	Y69307	Protein encoded by
38	37	86.0	928	21	Y69308	Protein encoded by
39	36	83.7	8	21	Y79127	Peptide antagonist
40	36	83.7	13	20	W86096	Peptide from human
41	36	83.7	13	20	W86096	Peptide from human
42	36	83.7	13	16	W86088	Peptide from human
43	36	83.7	18	16	W73953	Igg heavy chain fr
44	36	83.7	18	16	R82833	N-terminal of C179
45	36	83.7	20	20	W94487	Human adult heart

RESULT 1	
ID	Y79115 standard; Peptide; 8 AA.
Y79115	
AC	Y79115:
DT	05-JUN-2000 (first entry)
XX	Peptide antagonist of zonulin.
DE	
XX	Zonulin: antagonist; zonula occludens toxin receptor;
KW	blood-brain barrier; antiinflammatory; cerebroprotective;
KW	neuroprotective; dermatological; antitumor; antiviral;
KW	antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;
KW	hypotensive; immunosuppressive; antiparasitic; vasotropic;
KW	gastrointestinal inflammation; therapy.
OS	Synthetic.
XX	
PN	WO200007609-A1.
PD	17-FEB-2000.
XX	
PE	28-JUL-1999; 99WO-US16683.
XX	
PR	03-AUG-1998; 98US-0127815.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fasano A;
XX	
DR	WPI; 2000-205565/18.
XX	
PT	New peptide antagonist of zonulin useful as antiinflammatory agent for

PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis  
 XX  
 PS Claim 1, Page 44; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT  
 CC and human zonulins, which is believed to be critical for receptor  
 CC binding. They can be prepared by chemical synthesis or by use of  
 CC recombinant DNA techniques. The peptide antagonists are used as an  
 CC antiinflammatory agents in the treatment of gastrointestinal  
 CC inflammation, where they bind to the ZOT receptor in the intestine  
 CC and yet does not physiologically modulate the opening of TJ in the  
 CC intestine. Gastrointestinal inflammation conditions give rise to  
 CC increased intestinal permeability and the peptide is useful for  
 CC treating intestinal conditions that cause protein losing enteropathy  
 CC caused by infection, e.g. Clostridium difficile infection,  
 CC enterocolitis, shigellosis, viral gastroenteritis, parasite  
 CC infestation, bacterial overgrowth, whipple's disease, diseases with  
 CC mucosal erosion or ulcerations, e.g. gastritis, gastric cancer,  
 CC collagenous colitis, inflammatory bowel disease, diseases marked by  
 CC lymphatic obstruction, e.g. congenital intestinal lymphangiectasia,  
 CC sarcoidosis lymphoma, mesenteric tuberculosis, and after surgical  
 CC correction of congenital heart disease with Fontan's operation,  
 CC mucosal diseases without ulceration, e.g. Menetrier's disease,  
 CC coeliac disease, eosinophilic gastroenteritis, and immune diseases,  
 CC e.g. systemic lupus erythematosus or food allergies, primarily to  
 CC milk.  
 CC  
 SQ Sequence 8 AA;

Query Match 100.0%; Score 43; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVPG 8  
 |||||  
 Db 1 grglvpg 8

RESULT 2  
 R40924  
 ID R40924 standard; Protein; 246 AA.

AC R40924;

DT 17-FEB-1994 (first entry)

DE Protein able to bind to HIV-1 tat protein.

DE Antibodies: antigen binding proteins; library; HIV;

KW Human Immunodeficiency Virus.

OS Synthetic.

PN EP557897-A.

PD 01-SEP-1993.

PF 19-FEB-1993; 93EP-0102609.

PR 28-FEB-1992; 92US-0843125.

PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.

PI Dillon PJ, Rosen CA;

DR MPI: 1993-274375/35.

DR N-PSDB; Q48605.

XX Antigen-binding proteins and corresp. synthetic genes - are  
 PT constructed synthetically and used to make antibody library, in  
 PT disease diagnosis, etc.  
 XX  
 PS Claim 13; Page 22; 40pp; English.

CC The synthetic gene encodes a protein corresponding to an antibody  
 CC capable of binding to a specific antigen. In this case the HIV-1 tat  
 CC protein. Many synthetic genes are synthesised, each containing a  
 CC predetermined nucleotide region encoding the framework regions of  
 CC the heavy and light chains of antibody and underdetermined nucleotide  
 CC regions which are random sequences. The genes are then used in the  
 CC construction of vectors which are subsequently used to transform  
 CC microbes. The proteins thus produced are screened for binding  
 CC activity to the specific antigen.

SQ Sequence 246 AA;

Query Match 100.0%; Score 43; DB 14; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVPG 8  
 |||||  
 Db 8 grglvpg 15

RESULT 3  
 Y79113  
 ID Y79113 standard; Peptide; 8 AA.

AC Y79113;

DT 05-JUN-2000 (first entry)

DE Peptide antagonist of zonulin.

DE Zonulin: antagonist; zonula occludens toxin receptor;  
 KW blood-brain barrier; antiinflammatory; cerebroprotective;  
 KW neuroprotective; dermatological; antiulcer; antiviral;  
 KW antibacterial; cytostatic; anti-HIV; vulnery; antiallergic;  
 KW hypotensive; immunosuppressive; antiparasitic; vasotropic;  
 KW gastrointestinal inflammation; therapy.

OS Synthetic.

PN WO200007609-A1.

PD 17-FEB-2000.

PF 28-JUL-1999; 99WO-US16683.

PR 03-AUG-1998; 98US-0127815.

PA (UYMA-) UNIV MARYLAND BALTIMORE.

PI Fasano A;

DR WPI: 2000-205565/18.

PT New peptide antagonist of zonulin useful as antiinflammatory agent for  
 PT treating cerebral ischemia, stroke, cerebral edema, gastritis,  
 PT shigellosis, viral gastroenteritis, meningitis, encephalomyelitis -

PS Claim 1; Page 43; 69pp; English.

CC This present sequence is that of a peptide antagonist of zonulin  
 CC (2), one of 25 such peptides (see Y79105-29) of the invention,  
 CC which bind to a zonula occludens toxin (ZOT) receptor, yet do not  
 CC physiologically modulate the opening of mammalian tight junctions  
 CC (TJ). The peptide antagonists are based on a common motif of ZOT



CC peptides, as defined above but which have epitopes which show an  
 CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
 CC antiidiotypic antibody fragments against the MUC1-conformation epitope  
 CC having one of 31 approximately 60 residue amino acids sequences, all  
 CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
 CC having one of 16 9-17 residue amino acid sequences, all fully in the  
 CC specification; (5) antiidiotypic antibody fragments against the TF  
 CC antigen having one of 24 approximately 200 residue amino acid sequences,  
 CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
 CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
 CC the specification; and (7) DNA sequences encoding the fragments and  
 CC derivatives defined in (3), (4), (5), or (6) (V1) and (V2) are used to treat  
 CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
 CC and parasites. The vaccines are effective in cases where vaccination has  
 CC previously not been possible.

CC  
 XX Sequence 240 AA;

Query Match 88.4%; Score 38; DB 22; Length 240;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVQPG 8  
 | | | | |  
 Db 8 geglvpqg 15

RESULT 6  
 B45992  
 ID B45992 standard; peptide; 240 AA.  
 AC B45992;  
 XX  
 XX 23-MAR-2001 (first entry)  
 DE Human MUC-1 scfv clone Q1.3.  
 XX  
 XX MUC1: human; vaccine; conformation-dependent antigen; antibody; cancer;  
 KM antiidiotypic antibody; cytostatic; virucidal; antibacterial; TF antigen;  
 KM antiparasitic; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073430-A2.  
 PD 07-DEC-2000.  
 XX  
 XX 29-MAY-2000; 2000WO-DE01809.  
 PF  
 XX 27-MAY-1999; 99DE-1024405.  
 PR 09-SEP-1999; 99DE-1043016.  
 XX  
 XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 PA  
 PI Goletz S, Karsten U;  
 XX WPI: 2001-049937/06.  
 DR  
 XX  
 PT Vaccines against conformation-dependent or non-peptide antigens, based  
 PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
 PT antitumor vaccines  
 XX  
 XX  
 PS Disclosure: Page 5-9; 36pp; German.  
 CC This invention describes a novel vaccine (V1) against  
 CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
 CC antibody, or peptide which immunologically imitates CDA, is new. (I)  
 CC encodes a region of an antiidiotypic antibody (Ab2) or another peptide  
 CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
 CC or an antigen binding molecule; and (b) immunologically mimics the  
 CC initial antigen. The epitope is partially or completely  
 CC conformation-dependent, and has an immunogenic structure defined by a

CC specific spatial conformation of amino acids. (I) is used in the form  
 CC of linear or circular naked DNA and/or with a viral vector and/or  
 CC adjuvants. The products of the invention have cytostatic, virucidal,  
 CC antibacterial and antiparasitic. The invention also describes (1) a  
 CC corresponding vaccine (V2) against antigens which are not proteins or  
 CC peptides, as defined above but which have epitopes which show an  
 CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
 CC antiidiotypic antibody fragments against the MUC1-conformation epitope  
 CC having one of 31 approximately 60 residue amino acids sequences, all  
 CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
 CC having one of 16 9-17 residue amino acid sequences, all fully in the  
 CC specification; (5) antiidiotypic antibody fragments against the TF  
 CC antigen having one of 24 approximately 200 residue amino acid sequences,  
 CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
 CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
 CC the specification; and (7) DNA sequences encoding the fragments and  
 CC derivatives defined in (3), (4), (5), or (6) (V1) and (V2) are used to treat  
 CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
 CC and parasites. The vaccines are effective in cases where vaccination has  
 CC previously not been possible.

CC  
 XX Sequence 240 AA;

Query Match 88.4%; Score 38; DB 22; Length 240;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRLVQPG 8  
 | | | | |  
 Db 8 geglvpqg 15

RESULT 7  
 B45993  
 ID B45993 standard; peptide; 240 AA.  
 AC B45993;  
 XX  
 XX 23-MAR-2001 (first entry)  
 DE Human MUC-1 scfv clone Q1.2.  
 XX  
 XX MUC1: human; vaccine; conformation-dependent antigen; antibody; cancer;  
 KM antiidiotypic antibody; cytostatic; virucidal; antibacterial; TF antigen;  
 KM antiparasitic; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073430-A2.  
 PD 07-DEC-2000.  
 XX  
 XX 29-MAY-2000; 2000WO-DE01809.  
 PF  
 XX 27-MAY-1999; 99DE-1024405.  
 PR 09-SEP-1999; 99DE-1043016.  
 XX  
 XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 PA  
 PI Goletz S, Karsten U;  
 XX WPI: 2001-049937/06.  
 DR  
 XX  
 PT Vaccines against conformation-dependent or non-peptide antigens, based  
 PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
 PT antitumor vaccines  
 XX  
 XX  
 PS Disclosure: Page 5-9; 36pp; German.  
 CC This invention describes a novel vaccine (V1) against  
 CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
 CC antibody, or peptide which immunologically imitates CDA, is new. (I)

CC encodes a region of an antiidiotypic antibody (Ab2) or another peptide  
CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
CC or an antigen binding molecule; and (b) immunologically mimics the  
CC initial antigen. The epitope is partially or completely  
CC conformation-dependent, and has an immunogenic structure defined by a  
CC specific spatial conformation of amino acids. (1) is used in the form  
CC of linear or circular naked DNA and/or with a viral vector and/or  
CC adjuvants. The products of the invention have cytostatic, virucidal,  
CC antibacterial and antiparasitic. The invention also describes (1) a  
CC corresponding vaccine (V2) against antigens which are not proteins or  
CC peptides, as defined above but which have epitopes which show an  
CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
CC antiidiotypic antibody fragments against the MUC1-conformation epitope  
CC having one of 31 approximately 60 residue amino acids sequences, all  
CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
CC having one of 16-9-17 residue amino acid sequences, all fully in the  
CC specification; (5) antiidiotypic antibody fragments against the TF  
CC antigen having one of 24 approximately 200 residue amino acid sequences,  
CC fully defined in the specification; (6) TF carbohydrate epitope mimetics  
CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
CC the specification; and (7) DNA sequences encoding the fragments and  
CC derivatives defined in (3, 4, 5, or 6). (V1) and (V2) are used to treat  
CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
CC and parasites. The vaccines are effective in cases where vaccination has  
CC previously not been possible.

XX  
SQ Sequence 240 AA:

Query Match 88.4%; Score 38; DB 22; Length 240;  
Best Local Similarity 87.5%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVPG 8  
I I I I I I I I  
Db 8 geglvpg 15

RESULT 8  
B45994 ID B45994 standard; peptide: 240 AA.  
XX

AC B45994;

DT 23-MAR-2001 (first entry)

XX Human MUC-1 scfv clone Q4.

XX MUC1: human; vaccine: conformation-dependent antigen; antibody: cancer;

KM antiidiotypic antibody; cytostatic; virucidal; antibacterial; TF antigen;

KM antiparasitic; infectious disease.

XX Homo sapiens.

XX MO200073430-A2.

PN 07-DEC-2000.

XX 29-MAY-2000; 2000MO-DE01809.

XX 27-MAY-1999; 99DE-1024405.

PR 09-SEP-1999; 99DE-1043016.

XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

XX Goletz S, Karsten U;

XX WPI; 2001-049937/06.

PT Vaccines against conformation-dependent or non-peptide antigens, based  
PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
PT antitumor vaccines  
XX

PS Disclosure; Page 5-9; 36pp; German.

XX This invention describes a novel vaccine (V1) against  
CC conformation-dependent antigens (CDA) comprising DNA (1) and/or an  
CC antibody, or peptide which immunologically imitates CDA, is new. (1)  
CC encodes a region of an antiidiotypic antibody (Ab2) or another peptide  
CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
CC or an antigen binding molecule; and (b) immunologically mimics the  
CC initial antigen. The epitope is partially or completely  
CC conformation-dependent, and has an immunogenic structure defined by a  
CC specific spatial conformation of amino acids. (1) is used in the form  
CC of linear or circular naked DNA and/or with a viral vector and/or  
CC adjuvants. The products of the invention have cytostatic, virucidal,  
CC antibacterial and antiparasitic. The invention also describes (1) a  
CC corresponding vaccine (V2) against antigens which are not proteins or  
CC peptides, as defined above but which have epitopes which show an  
CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
CC antiidiotypic antibody fragments against the MUC1-conformation epitope  
CC having one of 31 approximately 60 residue amino acids sequences, all  
CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
CC having one of 16-9-17 residue amino acid sequences, all fully in the  
CC specification; (5) antiidiotypic antibody fragments against the TF  
CC antigen having one of 24 approximately 200 residue amino acid sequences,  
CC fully defined in the specification; (6) TF carbohydrate epitope mimetics  
CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
CC the specification; and (7) DNA sequences encoding the fragments and  
CC derivatives defined in (3, 4, 5, or 6). (V1) and (V2) are used to treat  
CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
CC and parasites. The vaccines are effective in cases where vaccination has  
CC previously not been possible.

XX  
SQ Sequence 240 AA:

Query Match 88.4%; Score 38; DB 22; Length 240;  
Best Local Similarity 87.5%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVPG 8  
I I I I I I I I  
Db 8 geglvpg 15

RESULT 9  
B45995 ID B45995 standard; peptide: 240 AA.  
XX

AC B45995;

DT 23-MAR-2001 (first entry)

XX Human MUC-1 scfv clone R2.

XX MUC1: human; vaccine: conformation-dependent antigen; antibody: cancer;

KM antiidiotypic antibody; cytostatic; virucidal; antibacterial; TF antigen;

KM antiparasitic; infectious disease.

XX Homo sapiens.

XX MO200073430-A2.

PN 07-DEC-2000.

XX 29-MAY-2000; 2000MO-DE01809.

XX 27-MAY-1999; 99DE-1024405.

PR 09-SEP-1999; 99DE-1043016.

XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

XX Goletz S, Karsten U;

XX WPI; 2001-049937/06.

XX Vaccines against conformation-dependent or non-peptide antigens, based  
 PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
 PT antitumor vaccines -  
 XX  
 PS Disclosure; Page 5-9; 36pp; German.

XX This invention describes a novel vaccine (V1) against  
 CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
 CC antibody, or peptide which immunologically imitates CDA, is new. (1)  
 CC encodes a region of an antidiotypic antibody (Ab2) or another peptide  
 CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
 CC or an antigen binding molecule; and (b) immunologically mimics the  
 CC initial antigen. The epitope is partially or completely  
 CC conformation-dependent, and has an immunogenic structure defined by a  
 CC specific spatial conformation of amino acids. (I) is used in the form  
 CC of linear or circular naked DNA and/or with a viral vector and/or  
 CC adjuvants. The products of the invention have cytostatic, virucidal,  
 CC antibacterial and antiparasitic. The invention also describes (1) a  
 CC corresponding vaccine (V2) against antigens which are not proteins or  
 CC peptides, as defined above but which have epitopes which show an  
 CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
 CC antidiotypic antibody fragments against the MUC1-conformation epitope  
 CC having one of 31 approximately 60 residue amino acids sequences, all  
 CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
 CC having one of 16 9-17 residue amino acid sequences, all fully in the  
 CC specification; (5) antidiotypic antibody fragments against the TF  
 CC antigen having one of 24 approximately 200 residue amino acid sequences,  
 CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
 CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
 CC the specification; and (7) DNA sequences encoding the fragments and  
 CC derivatives defined in (3, 4, 5, or 6). (V1) and (V2) are used to treat  
 CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
 CC and parasites. The vaccines are effective in cases where vaccination has  
 CC previously not been possible.  
 XX  
 CC Sequence 240 AA:  
 SQ

Query Match 88.4%; Score 38; DB 22; Length 240;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQPG 8  
 I | | | | | | |  
 Db 8 geglvqpg 15

RESULT 10  
 B45996  
 ID B45996 standard; Peptide: 240 AA.

AC B45996;  
 XX  
 DT 23-MAR-2001 (first entry)  
 XX  
 DE Human MUC-1 scFv clone Q15.  
 XX  
 XX MUC1; human; vaccine; conformation-dependent antigen; antibody; cancer;  
 KW antidiotypic antibody; cytostatic; virucidal; antibacterial; TF antigen;  
 KW antiparasitic; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073430-A2.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 29-MAY-2000; 2000WO-DE01809.  
 XX  
 PR 27-MAY-1999; 99DE-1024405.  
 PR 09-SEP-1999; 99DE-1043016.  
 XX

PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 XX  
 XX Goletz S, Karsten U;  
 XX  
 DR WPI; 2001-049937/06.  
 XX  
 PT Vaccines against conformation-dependent or non-peptide antigens, based  
 PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
 PT antitumor vaccines -  
 XX  
 PS Disclosure; Page 5-9; 36pp; German.

XX This invention describes a novel vaccine (V1) against  
 CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
 CC antibody, or peptide which immunologically imitates CDA, is new. (1)  
 CC encodes a region of an antidiotypic antibody (Ab2) or another peptide  
 CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
 CC or an antigen binding molecule; and (b) immunologically mimics the  
 CC initial antigen. The epitope is partially or completely  
 CC conformation-dependent, and has an immunogenic structure defined by a  
 CC specific spatial conformation of amino acids. (I) is used in the form  
 CC of linear or circular naked DNA and/or with a viral vector and/or  
 CC adjuvants. The products of the invention have cytostatic, virucidal,  
 CC antibacterial and antiparasitic. The invention also describes (1) a  
 CC corresponding vaccine (V2) against antigens which are not proteins or  
 CC peptides, as defined above but which have epitopes which show an  
 CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
 CC antidiotypic antibody fragments against the MUC1-conformation epitope  
 CC having one of 31 approximately 60 residue amino acids sequences, all  
 CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
 CC having one of 16 9-17 residue amino acid sequences, all fully in the  
 CC specification; (5) antidiotypic antibody fragments against the TF  
 CC antigen having one of 24 approximately 200 residue amino acid sequences,  
 CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
 CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
 CC the specification; and (7) DNA sequences encoding the fragments and  
 CC derivatives defined in (3, 4, 5, or 6). (V1) and (V2) are used to treat  
 CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
 CC and parasites. The vaccines are effective in cases where vaccination has  
 CC previously not been possible.  
 XX  
 CC Sequence 240 AA:  
 SQ

Query Match 88.4%; Score 38; DB 22; Length 240;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQPG 8  
 I | | | | | | |  
 Db 8 geglvqpg 15

RESULT 11  
 B45997  
 ID B45997 standard; Peptide: 240 AA.

AC B45997;  
 XX  
 DT 23-MAR-2001 (first entry)  
 XX  
 DE Human MUC-1 scFv clone R10.  
 XX  
 XX MUC1; human; vaccine; conformation-dependent antigen; antibody; cancer;  
 KW antidiotypic antibody; cytostatic; virucidal; antibacterial; TF antigen;  
 KW antiparasitic; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073430-A2.  
 XX  
 PD 07-DEC-2000.  
 XX

PE 29-MAY-2000; 2000MO-DE01809.  
XX  
XX 27-MAY-1999; 99DE-1024405.  
PR 09-SEP-1999; 99DE-1043016.  
XX  
XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
XX  
XX Goletz S, Karsten U;  
XX WPI; 2001-049937/06.  
XX  
XX Vaccines against conformation-dependent or non-peptide antigens, based  
PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
PT antitumor vaccines  
PS  
PS Disclosure; Page 5-9; 36pp; German.  
XX  
XX This invention describes a novel vaccine (V1) against  
CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
CC antibody, or peptide which immunologically imitates CDA, is new. (I)  
CC encodes a region of an antidiotytic antibody (Ab2) or another peptide  
CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
CC or an antigen binding molecule; and (b) immunologically mimics the  
CC initial antigen. The epitope is partially or completely  
CC conformation-dependent, and has an immunogenic structure defined by a  
CC specific spatial conformation of amino acids. (I) is used in the form  
CC of linear or circular naked DNA and/or with a viral vector and/or  
CC adjuvants. The products of the invention have cytostatic, virucidal,  
CC antibacterial and antiparasitic. The invention also describes (1) a  
CC corresponding vaccine (V2) against antigens which are not proteins or  
CC peptides, as defined above but which have epitopes which show an  
CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
CC antidiotytic antibody fragments against the MUC1-conformation epitope  
CC having one of 31 approximately 60 residue amino acid sequences, all  
CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
CC having one of 16 9-17 residue amino acid sequences, all fully in the  
CC specification; (5) antidiotytic antibody fragments against the TF  
CC antigen having one of 24 approximately 200 residue amino acid sequences,  
CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
CC the specification; and (7) DNA sequences encoding the fragments and  
CC derivatives defined in (3), 4, 5, or 6). (V1) and (V2) are used to treat  
CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
CC and parasites. The vaccines are effective in cases where vaccination has  
CC previously not been possible.  
XX  
XX Sequence 240 AA:  
SQ  
Query Match 88.4%; Score 38; DB 22; Length 240;  
Best Local Similarity 87.5%; Pred. NO. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GRGLVPG 8  
Db 8 geglvpg 15  
RESULT 12  
B45998 B45998 standard; Peptide: 240 AA.  
XX  
XX B45998;  
AC  
XX 23-MAR-2001 (first entry)  
DT XX Human MUC-1 scFv clone C5.  
DE  
XX  
XX MUC1: human; vaccine; conformation-dependent antigen; antibody; cancer;  
KM antidiotytic antibody; cytostatic; virucidal; antibacterial; TF antigen;  
KW antiparasitic; infectious disease.  
XX  
XX Homo sapiens.  
OS

XX  
XX WO200073430-A2.  
PN  
XX  
XX 07-DEC-2000.  
PD  
XX  
XX 29-MAY-2000; 2000MO-DE01809.  
PE  
XX  
XX 27-MAY-1999; 99DE-1024405.  
PR 09-SEP-1999; 99DE-1043016.  
XX  
XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
XX  
XX Goletz S, Karsten U;  
XX WPI; 2001-049937/06.  
XX  
XX Vaccines against conformation-dependent or non-peptide antigens, based  
PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
PT antitumor vaccines  
PS  
PS Disclosure; Page 5-9; 36pp; German.  
XX  
XX This invention describes a novel vaccine (V1) against  
CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
CC antibody, or peptide which immunologically imitates CDA, is new. (I)  
CC encodes a region of an antidiotytic antibody (Ab2) or another peptide  
CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
CC or an antigen binding molecule; and (b) immunologically mimics the  
CC initial antigen. The epitope is partially or completely  
CC conformation-dependent, and has an immunogenic structure defined by a  
CC specific spatial conformation of amino acids. (I) is used in the form  
CC of linear or circular naked DNA and/or with a viral vector and/or  
CC adjuvants. The products of the invention have cytostatic, virucidal,  
CC antibacterial and antiparasitic. The invention also describes (1) a  
CC corresponding vaccine (V2) against antigens which are not proteins or  
CC peptides, as defined above but which have epitopes which show an  
CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
CC antidiotytic antibody fragments against the MUC1-conformation epitope  
CC having one of 31 approximately 60 residue amino acid sequences, all  
CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
CC having one of 16 9-17 residue amino acid sequences, all fully in the  
CC specification; (5) antidiotytic antibody fragments against the TF  
CC antigen having one of 24 approximately 200 residue amino acid sequences,  
CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
CC having one of 25 7-13 residue amino acid sequences, all fully defined in  
CC the specification; and (7) DNA sequences encoding the fragments and  
CC derivatives defined in (3), 4, 5, or 6). (V1) and (V2) are used to treat  
CC cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria  
CC and parasites. The vaccines are effective in cases where vaccination has  
CC previously not been possible.  
XX  
XX Sequence 240 AA:  
SQ  
Query Match 88.4%; Score 38; DB 22; Length 240;  
Best Local Similarity 87.5%; Pred. NO. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GRGLVPG 8  
Db 8 geglvpg 15  
RESULT 13  
B45999 B45999 standard; Peptide: 240 AA.  
XX  
XX B45999;  
AC  
XX 23-MAR-2001 (first entry)  
DT XX Human MUC-1 scFv clone N6.  
DE  
XX

OS	Homo sapiens.
XX	
XX	MO200073430-A2.
PB	07-DEC-2000.
XX	
PF	29-MAY-2000; 2000WO-DE01809.
XX	
PR	27-MAY-1999; 99DE-1024405.
PR	09-SEP-1999; 99DE-1043016.
XX	
PA	(DELBA-) DELBRUECK CENT MOLEKULARE MEDIZIN.MAX.
XX	
PI	Goletz S, Karsten U;
XX	
DR	WPI; 2001-049937/06.
XX	
PT	Vaccines against conformation-dependent or non-peptide antigens, based
PT	on DNA encoding peptide which mimics the antigen, useful e.g. as
PT	antitumor vaccines -
XX	
PS	Disclosure: Page 5-9; 36pp; German.
XX	
CC	This invention describes a novel vaccine (V1) against
CC	conformation-dependent antigens (CDA) comprising DNA (I) and/or an
CC	antibody, or peptide which immunologically imitates CDA, is new. (I)
CC	encodes a region of an antidiotypic antibody (Ab2) or another peptide
CC	which: (a) specifically binds to the binding site of an antibody (Ab1)
CC	or an antigen binding molecule; and (b) immunologically mimics the
CC	initial antigen. The epitope is partially or completely
CC	conformation-dependent, and has an immunogenic structure defined by a
CC	specific spatial conformation of amino acids. (I) is used in the form
CC	of linear or circular naked DNA and/or with a viral vector and/or
CC	adjuvants. The products of the invention have cytostatic, virucidal,
CC	antibacterial and antiparasitic. The invention also describes (1) a
CC	corresponding vaccine (V2) against antigens which are not proteins or
CC	peptides, as defined above but which have epitopes which show an
CC	immunogenic structure; (2) preparing (V1) and (V2); (3) human
CC	antidiotypic antibody fragments against the MUC1-conformation epitope
CC	having one of 31 approximately 60 residue amino acids sequences, all
CC	fully defined in the specification; (4) MUC1-conformation epitope mimics
CC	having one of 16 9-17 residue amino acid sequences, all fully in the
CC	specification; (5) antidiotypic antibody fragments against the TF
CC	antigen having one of 24 approximately 200 residue amino acid sequences,
CC	fully defined in the specification; (6) TF carbohydrate epitope mimetics
CC	having one of 25 7-13 residue amino acid sequences, all fully defined in
CC	the specification; and (7) DNA sequences encoding the fragments and
CC	derivatives defined in (3), (4), (5), or (6). (V1) and (V2) are used to treat
CC	cancer, and infectious diseases, e.g. caused by prions, viruses, bacteria
CC	and parasites. The vaccines are effective in cases where vaccination has
CC	previously not been possible.
XX	
SQ	Sequence      240 AA:
	Query Match                 88.4%;   Score 38;   DB 22;   Length 240;
	Best Local Similarity       87.5%;   Pred. No. 17;
	Matches       7;   Conservative       0;   Mismatches       1;   Indels       0;   Gaps       0
DB	
	1 GRGLWPG 8
	8 geglwpqg 15
RESULT	14
ID	B46000 standard; Peptide; 240 AA.
XX	
AC	B46000;

[illegible]



RESULT 15  
 B46001  
 ID B46001 standard; Peptide: 240 AA.  
 AC B46001;  
 XX  
 DT 23-MAR-2001 (first entry)  
 XX  
 DE Human MUC-1 scFv clone R6.  
 XX  
 KW MUC1; human; vaccine; conformation-dependent antigen; antibody; cancer;  
 KW antidiotytic antibody; cytostatic; virucidal; antibacterial; TF antigen;  
 KW antiparasitic; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN M0200073430-A2.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 29-MAY-2000; 2000MO-DE01809.  
 XX  
 PR 27-MAY-1999; 99DE-1024405.  
 PR 09-SEP-1999; 99DE-1043016.  
 XX  
 PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 XX  
 PI Goletz S, Karsten U;  
 XX  
 DR WPI: 2001-049937/06.  
 XX  
 PT Vaccines against conformation-dependent or non-peptide antigens, based  
 PT on DNA encoding peptide which mimics the antigen, useful e.g. as  
 PT antitumor vaccines  
 XX  
 PS Disclosure: Page 5-9; 36pp; German.  
 XX  
 CC This invention describes a novel vaccine (V1) against  
 CC conformation-dependent antigens (CDA) comprising DNA (I) and/or an  
 CC antibody, or peptide which immunologically imitates CDA, is new. (I)  
 CC encodes a region of an antidiotytic antibody (Ab2) or another peptide  
 CC which: (a) specifically binds to the binding site of an antibody (Ab1)  
 CC or an antigen binding molecule; and (b) immunologically mimics the  
 CC initial antigen. The epitope is partially or completely  
 CC conformation-dependent, and has an immunogenic structure defined by a  
 CC specific spatial conformation of amino acids. (I) is used in the form  
 CC of linear or circular naked DNA and/or with a viral vector and/or  
 CC adjuvants. The products of the invention have cytostatic, virucidal,  
 CC antibacterial and antiparasitic. The invention also describes (1) a  
 CC corresponding vaccine (V2) against antigens which are not proteins or  
 CC peptides, as defined above but which have epitopes which show an  
 CC immunogenic structure; (2) preparing (V1) and (V2); (3) human  
 CC antidiotytic antibody fragments against the MUC1-conformation epitope  
 CC having one of 31 approximately 60 residue amino acid sequences, all  
 CC fully defined in the specification; (4) MUC1-conformation epitope mimics  
 CC having one of 16-9-17 residue amino acid sequences, all fully in the  
 CC specification; (5) antidiotytic antibody fragments against the TF  
 CC antigen having one of 24 approximately 200 residue amino acid sequences,  
 CC fully defined in the specification; (6) TF carbohydrate epitope mimics  
 CC having one of 25-7-13 residue amino acid sequences, all fully defined in  
 CC the specification; and (7) DNA sequences encoding the fragments and  
 CC derivatives defined in (3, 4, 5, or 6). (V1) and (V2) are used to treat  
 CC cancer, and infectious diseases, e.g. caused by pitons, viruses, bacteria  
 CC and parasites. The vaccines are effective in cases where vaccination has  
 CC previously not been possible.  
 XX  
 SO Sequence 240 AA;

OY 1 GRLVOPG 8  
 1 1 1 1 1 1 1  
 Db 8 geglvpqg 15

Search completed: June 13, 2001, 14:14:35  
 Job time: 376 sec

Query Match 88.4%; Score 38; DB 22; Length 240;  
 Best Local Similarity 87.5%; Pred. NO. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2001, 14:10:43 ; Search time 130.61 Seconds

(without alignments)  
4.209 Million cell updates/sec

Title: PCT-US01-05825A-11

Perfect score: 43

Sequence: 1 GRGLVQPG 8

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

PIR\_67:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	37	86.0	119	2 S24513	Ig heavy chain V r
2	36	83.7	36	2 D31485	Ig heavy chain V r
3	36	83.7	38	2 S33402	Ig heavy chain V r
4	36	83.7	59	2 A27606	Ig heavy chain V-a
5	36	83.7	59	2 S36381	Ig heavy chain V r
6	36	83.7	60	2 S36382	Ig heavy chain V r
7	36	83.7	82	2 C34964	Ig heavy chain pre
8	36	83.7	94	2 D35913	Ig heavy chain V r
9	36	83.7	94	2 I67528	CD33 antigen homol
10	36	83.7	95	2 I67527	Ig heavy chain V r
11	36	83.7	97	2 PH0872	Ig heavy chain V r
12	36	83.7	97	2 S26885	Ig heavy chain V r
13	36	83.7	97	2 S26886	Ig heavy chain V r
14	36	83.7	97	2 S26890	Ig heavy chain V r
15	36	83.7	97	2 S46462	Ig heavy chain V r
16	36	83.7	97	2 S26895	Ig heavy chain V r
17	36	83.7	97	2 S24855	Ig heavy chain V r
18	36	83.7	98	2 PL0121	Ig heavy chain V-I
19	36	83.7	98	2 PH0874	Ig heavy chain V-I
20	36	83.7	98	2 PL0123	Ig heavy chain V-I
21	36	83.7	98	2 S26896	Ig heavy chain V r
22	36	83.7	98	2 S29545	Ig heavy chain V r
23	36	83.7	98	2 S26927	Ig heavy chain V r
24	36	83.7	98	2 S26932	Ig heavy chain V r
25	36	83.7	98	2 S26891	Ig heavy chain V r
26	36	83.7	98	2 S26894	Ig heavy chain V r
27	36	83.7	98	2 S26889	Ig heavy chain V r
28	36	83.7	98	2 S26933	Ig heavy chain V r
29	36	83.7	98	2 S26934	Ig heavy chain V r

30	36	83.7	100	2 PL0122	Ig heavy chain V-I
31	36	83.7	100	2 S66896	Ig heavy chain V r
32	36	83.7	100	2 S26925	Ig heavy chain V r
33	36	83.7	100	2 S26926	Ig heavy chain V r
34	36	83.7	101	2 PU0003	Ig heavy chain V r
35	36	83.7	101	2 PU0002	Ig heavy chain V r
36	36	83.7	105	2 S38488	Ig heavy chain - h
37	36	83.7	105	2 PL0255	Ig heavy chain V r
38	36	83.7	106	2 S24521	Ig heavy chain V r
39	36	83.7	111	1 MHMS76	Ig heavy chain V-I
40	36	83.7	111	2 PH1028	Ig heavy chain V r
41	36	83.7	111	2 S69911	Ig V-D-J region (R
42	36	83.7	111	2 S13693	Ig heavy chain V r
43	36	83.7	112	2 S13690	Ig heavy chain V r
44	36	83.7	113	1 AVMSAB	Ig heavy chain V-I
45	36	83.7	113	1 AVMS61	Ig heavy chain V-I

## ALIGNMENTS

RESULT 1  
S24513  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
R:Kaartinen, M.  
submitted to the EMBL Data Library, October 1991  
A:Reference number: S24490  
A:Accession: S24513  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-119 <KAA>  
A:Cross-references: EMBL:X66655; NID:951229; PIDN:CAA47217.1; PID:951230  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:34-119/Domain: immunoglobulin homology <1MM>

Query Match 86.0%; Score 37; DB 2; Length 119;  
Best Local Similarity 87.5%; Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQPG 8  
DB 27 GXGLVQPG 34

RESULT 2  
D31485  
Ig heavy chain V regions (3-24 and 5-14) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 30-May-1997  
R:Bedzyk, W.D.; Johnson, L.S.; Riordan, G.S.; Voss Jr., E.W.  
J. Biol. Chem. 264, 1565-1569, 1989  
A:Title: Comparison of variable region primary structures within an anti-fluorescein  
A:Reference number: A31485; MUID:89109167  
A:Accession: D31485  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-36 <BED>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 83.7%; Score 36; DB 2; Length 36;  
Best Local Similarity 87.5%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQPG 8  
I | | | | | | |

Db 8 GGGIVQPG 15

# RESULT 3

Ig heavy chain V region - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 17-Jul-1998  
 C:Accession: S33402; S36385  
 R:Kettleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
 Eur. J. Immunol. 23, 206-211, 1993  
 A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
 A:Reference number: S33391; MUID:93122092  
 A:Accession: S33402  
 A:Molecule type: mRNA  
 A:Residues: 1-38 <KET>  
 A:Cross-references: EMBL:X73019  
 A:Experimental source: strain BALB/c  
 R:Ansell, K.H.  
 Submitted to the EMBL Data Library, April 1993  
 A:Reference number: S36376  
 A:Accession: S36385  
 A:Molecule type: mRNA  
 A:Residues: 1-30 <ANS>  
 A:Cross-references: EMBL:X73019  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:1-38/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 83.7%; Score 36; DB 2; Length 38;  
 Best Local Similarity 87.5%; Pred. No. 1.4;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGIVQPG 8  
 | | | | | | |  
 Db 6 GGGIVQPG 13

# RESULT 4

A27606  
 Ig heavy chain V-a region (p26.9p2) - rabbit (fragments)  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 08-Mar-1989 #sequence\_revision 30-Jun-1991 #text\_change 16-Aug-1996  
 C:Accession: A27606  
 R:Currier, S.J.; Gallarda, J.L.; Knight, K.L.  
 J. Immunol. 140, 1651-1659, 1988  
 A:Title: Partial molecular genetic map of the rabbit V-H chromosomal region.  
 A:Reference number: A27606; MUID:88154464  
 A:Accession: A27606  
 A:Molecule type: DNA  
 A:Residues: 1-59 <CUR>  
 A:Cross-references: GB:M19706  
 A:Note: the authors translated the codon TCC for residue 28 as Asp  
 C:Genetics:  
 A:Introns: 30/3  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin

Query Match 83.7%; Score 36; DB 2; Length 59;  
 Best Local Similarity 87.5%; Pred. No. 2.1;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGIVQPG 8  
 | | | | | | |  
 Db 8 GGGIVQPG 15

# RESULT 5

S36381  
 Ig heavy chain V region - mouse (fragment)  
 C:Species: Mus musculus (house mouse)

C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 23-Jul-1999  
 C:Accession: S36381; S33396  
 R:Ansell, K.H.

Submitted to the EMBL Data Library, April 1993  
 A:Reference number: S36376  
 A:Accession: S36381  
 A:Molecule type: mRNA  
 A:Residues: 1-59 <ANS>  
 A:Cross-references: EMBL:X73012; NID:g295879; PIDN:CA51498.1; PID:g9339941  
 R:Kettleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
 Eur. J. Immunol. 23, 206-211, 1993  
 A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
 A:Reference number: S33391; MUID:93122092  
 A:Accession: S33396  
 A:Molecule type: mRNA  
 A:Residues: 3-43 <KET>  
 A:Cross-references: EMBL:X73012  
 A:Experimental source: strain BALB/c  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:14-59/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 83.7%; Score 36; DB 2; Length 59;  
 Best Local Similarity 87.5%; Pred. No. 2.1;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGIVQPG 8  
 | | | | | | |  
 Db 10 GGGIVQPG 17

# RESULT 6

S36382  
 Ig heavy chain V region - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 23-Jul-1999  
 C:Accession: S36382; S33397  
 R:Ansell, K.H.  
 Submitted to the EMBL Data Library, April 1993  
 A:Reference number: S36376  
 A:Accession: S36382  
 A:Molecule type: mRNA  
 A:Residues: 1-60 <ANS>  
 A:Cross-references: EMBL:X73011; NID:g295878; PIDN:CA51497.1; PID:g9339940  
 R:Kettleborough, C.A.; Saldanha, J.; Ansell, K.H.; Bendig, M.M.  
 Eur. J. Immunol. 23, 206-211, 1993  
 A:Title: Optimization of primers for cloning libraries of mouse immunoglobulin genes  
 A:Reference number: S33391; MUID:93122092  
 A:Accession: S33397  
 A:Molecule type: mRNA  
 A:Residues: 1-33 <KET>  
 A:Cross-references: EMBL:X73011  
 A:Experimental source: strain BALB/c  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:8-60/Domain: immunoglobulin homology (fragment) <IMM>

Query Match 83.7%; Score 36; DB 2; Length 60;  
 Best Local Similarity 87.5%; Pred. No. 2.1;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGIVQPG 8  
 | | | | | | |  
 Db 1 GGGIVQPG 8

# RESULT 7

C34964  
 Ig heavy chain precursor V-III region (Ab21) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-May-1997

C:Accession: C34964  
R:Sanz, I.; Casali, P.; Thomas, J.W.; Nockins, A.L.; Capra, J.D.  
J: Immunol. 142, 4054-4061, 1989  
A:Title: Nucleotide sequences of eight human natural antibody V-H regions reveals at  
A:Reference number: A92850; MUID:89235232  
A:Accession: C34964  
A:Molecule type: mRNA  
A:Residues: 1-82 <SAN>  
A:Cross-references: GB:M26996  
A:Note: the sequence shown differs from the authors' translation after residue 56  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 83.7%; Score 36; DB 2; Length 82;  
Best Local Similarity 87.5%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQPG 8  
| | | | |  
Db 27 GGGLVQPG 34

RESULT 8  
D25913  
Ig heavy chain V region (BFL) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 16-Aug-1996  
C:Accession: D25913  
R:Lawler, A.M.; Lin, P.S.; Gearhart, P.J.  
Proc. Natl. Acad. Sci. U.S.A. 84, 2454-2458, 1987  
A:Title: Adult B-cell repertoire is biased toward two heavy-chain variable-region genes  
A:Reference number: A94148; MUID:87175692  
A:Accession: D25913  
A:Molecule type: DNA  
A:Residues: 1-94 <LAN>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:11-94/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 94;  
Best Local Similarity 87.5%; Pred. No. 3.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQPG 8  
| | | | |  
Db 4 GGGLVQPG 11

RESULT 9  
167528  
CD3 antigen homolog - mouse (fragment)  
C:Species: Mus sp. (mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jan-2000  
C:Accession: 167528  
R:Chies, J.A.; Lembezat, M.P.; Freltas, A.A.  
Eur. J. Immunol. 24, 1657-1664, 1994  
A:Title: Entry of B lymphocytes into the persistent cell pool in non-immunized mice is r  
A:Reference number: 153392; MUID:94298870  
A:Accession: 167528  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-94 <RES>  
A:Cross-references: GB:S71357; NID:9550040  
C:Genetics:  
A:Gene: Ig VHx24b  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:11-94/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 94;  
Best Local Similarity 87.5%; Pred. No. 3.4;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GRGLVQPG 8  
| | | | |  
Db 4 GGGLVQPG 11

RESULT 10  
167527  
CD3 antigen homolog - mouse (fragment)  
C:Species: Mus sp. (mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jan-2000  
C:Accession: 167527  
R:Chies, J.A.; Lembezat, M.P.; Freltas, A.A.  
Eur. J. Immunol. 24, 1657-1664, 1994  
A:Title: Entry of B lymphocytes into the persistent cell pool in non-immunized mice 1  
A:Reference number: 153392; MUID:94298870  
A:Accession: 167527  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-95 <RES>  
A:Cross-references: GB:S71351; NID:9550039  
C:Genetics:  
A:Gene: Ig VHx24a  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:12-95/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 95;  
Best Local Similarity 87.5%; Pred. No. 3.4;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQPG 8  
| | | | |  
Db 5 GGGLVQPG 12

RESULT 11  
PH0872  
Ig heavy chain V region (anti-DNA, III-3R) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 16-Aug-1996  
C:Accession: PH0872  
R:Manheimer-Lory, A.; Katz, J.B.; Pillingner, M.; Ghossein, C.; Smith, A.; Diamond, B.  
J. Exp. Med. 174, 1639-1652, 1991  
A:Title: Molecular characteristics of antibodies bearing an anti-DNA-associated idiot  
A:Reference number: PH0862; MUID:92078875  
A:Accession: PH0872  
A:Molecule type: DNA  
A:Residues: 1-97 <MAN>  
A:Note: a "G" was inserted at the position of between 72nd and 73rd of the DNA sequen  
C:Comment: This antibody is produced by Epstein-Barr virus-transformed B cell that be  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:14-97/Domain: immunoglobulin homology <IMM>  
F:30-35/Region: complementarity-determining 1  
F:49-66/Region: complementarity-determining 2

Query Match 83.7%; Score 36; DB 2; Length 97;  
Best Local Similarity 87.5%; Pred. No. 3.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVQPG 8  
| | | | |  
Db 7 GGGLVQPG 14

RESULT 12  
S26885  
Ig heavy chain V region (DP-44) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 23-Jul-1999

C:Accession: S26885; S36596  
 R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Llewellyn, M.B.; Winter, G.  
 J. Mol. Biol. 227, 776-798, 1992  
 A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups of V  
 A:Reference number: S26885; MUID:93021117  
 A:Accession: S26885  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-97 <TOM1>  
 A:Cross-references: EMBL:212344  
 R:Tomlinson, M.  
 submitted to the EMBL Data Library, June 1992  
 A:Reference number: S36596  
 A:Accession: S36596  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-5, 'Q', 7-12, 'H', 14-97 <TOM2>  
 A:Cross-references: EMBL:212344; NID:932908; PIDN:CAA78214.1; PID:932909  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 97;  
 Best Local Similarity 87.5%; Pred. No. 3.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVOPG 8  
 | | | | |  
 Db 8 GGGLVOPG 15

RESULT 13  
 S26886  
 Ig heavy chain V region (DP-45) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
 C:Accession: S26886  
 R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Llewellyn, M.B.; Winter, G.  
 J. Mol. Biol. 227, 776-798, 1992  
 A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups of V  
 A:Reference number: S26885; MUID:93021117  
 A:Accession: S26886  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-97 <TOM>  
 A:Cross-references: EMBL:212345; NID:932910; PIDN:CAA78215.1; PID:932911  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 97;  
 Best Local Similarity 87.5%; Pred. No. 3.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVOPG 8  
 | | | | |  
 Db 8 GGGLVOPG 15

RESULT 14  
 S26890  
 Ig heavy chain V region (DP-48) - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
 C:Accession: S26890  
 R:Tomlinson, I.M.; Walter, G.; Marks, J.D.; Llewellyn, M.B.; Winter, G.  
 J. Mol. Biol. 227, 776-798, 1992  
 A:Title: The repertoire of human germline V(H) sequences reveals about fifty groups of V  
 A:Reference number: S26885; MUID:93021117  
 A:Accession: S26890  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-97 <TOM>  
 A:Cross-references: EMBL:212348; NID:932916; PIDN:CAA78218.1; PID:932917  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 97;  
 Best Local Similarity 87.5%; Pred. No. 3.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVOPG 8  
 | | | | |  
 Db 8 GGGLVOPG 15

RESULT 15  
 S46462  
 Ig heavy chain V region (YAC-5) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
 C:Accession: S46462  
 R:Cook, G.P.; Tomlinson, I.M.; Walter, G.; Riethman, H.; Carter, N.P.; Bulwela, L.;  
 Nature Genet. 7, 162-168, 1994  
 A:Title: A map of the human immunoglobulin V(H) locus completed by analysis of the te  
 A:Reference number: S46460; MUID:95004581  
 A:Accession: S46462  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-97 <COO>  
 A:Cross-references: EMBL:227504; NID:9505430; PIDN:CAA81824.1; PID:9505431  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 83.7%; Score 36; DB 2; Length 97;  
 Best Local Similarity 87.5%; Pred. No. 3.5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVOPG 8  
 | | | | |  
 Db 8 GGGLVOPG 15

Search completed: June 13, 2001, 14:10:43  
 Job time: 144 sec



*This Page Blank (uspto)*



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 13, 2001, 14:21:44 ; Search time 74.44 Seconds  
(without alignments)  
3.681 Million cell updates/sec

Title: PCT-US01-05825a-11

Perfect score: 43

Sequence: 1 GRGLVPG 8

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	83.7	111	1 HV35_MOUSE	P01804 mus musculu
2	36	83.7	113	1 HV27_MOUSE	P01796 mus musculu
3	36	83.7	113	1 HV28_MOUSE	P01797 mus musculu
4	36	83.7	113	1 HV29_MOUSE	P01798 mus musculu
5	36	83.7	113	1 HV30_MOUSE	P01799 mus musculu
6	36	83.7	113	1 HV31_MOUSE	P01800 mus musculu
7	36	83.7	113	1 HV34_MOUSE	P01803 mus musculu
8	36	83.7	115	1 HV32_MOUSE	P01801 mus musculu
9	36	83.7	115	1 HV33_MOUSE	P01802 mus musculu
10	36	83.7	115	1 HV36_MOUSE	P01806 mus musculu
11	36	83.7	116	1 HV3R_MOUSE	P01779 homo sapien
12	36	83.7	116	1 HV3R_MOUSE	P01786 mus musculu
13	36	83.7	117	1 HV17_MOUSE	P01764 homo sapien
14	36	83.7	117	1 HV3C_MOUSE	P01776 homo sapien
15	36	83.7	117	1 HV30_MOUSE	P01811 mus musculu
16	36	83.7	117	1 HV41_MOUSE	P18525 mus musculu
17	36	83.7	117	1 HV54_MOUSE	P01809 mus musculu
18	36	83.7	118	1 HV39_MOUSE	P80419 homo sapien
19	36	83.7	118	1 HV3V_MOUSE	P01807 mus musculu
20	36	83.7	119	1 HV37_MOUSE	P01808 mus musculu
21	36	83.7	119	1 HV38_MOUSE	P01774 homo sapien
22	36	83.7	119	1 HV3M_MOUSE	P01775 homo sapien
23	36	83.7	119	1 HV3N_MOUSE	P01777 homo sapien
24	36	83.7	119	1 HV3P_MOUSE	P01776 homo sapien
25	36	83.7	119	1 HV40_MOUSE	P01810 mus musculu
26	36	83.7	120	1 HV3E_MOUSE	P01789 mus musculu
27	36	83.7	122	1 HV20_MOUSE	P01790 mus musculu
28	36	83.7	122	1 HV21_MOUSE	P01787 mus musculu
29	36	83.7	123	1 HV18_MOUSE	P01788 mus musculu
30	36	83.7	123	1 HV19_MOUSE	P01791 mus musculu
31	36	83.7	123	1 HV22_MOUSE	P01792 mus musculu
32	36	83.7	123	1 HV23_MOUSE	P01793 mus musculu
33	36	83.7	123	1 HV24_MOUSE	

34	36	83.7	123	1 HV25_MOUSE	P01794 mus musculu
35	36	83.7	136	1 HV16_MOUSE	P01783 mus musculu
36	36	83.7	142	1 HV01_RAT	P01805 rattus norv
37	36	83.7	144	1 HV26_MOUSE	P01795 mus musculu
38	35	81.4	115	1 HV3E_MOUSE	P01776 homo sapien
39	35	81.4	116	1 HV05_CARAU	P19181 carassius a
40	35	81.4	1122	1 TERT_MOUSE	O70372 mus musculu
41	34	79.1	713	1 GAC1_HUMAN	O75325 homo sapien
42	33	76.7	114	1 HV3B_HUMAN	P01763 homo sapien
43	33	76.7	119	1 HV3I_HUMAN	P01770 homo sapien
44	33	76.7	121	1 HV3J_HUMAN	P01771 homo sapien
45	33	76.7	126	1 HV3K_HUMAN	P01772 homo sapien

## ALIGNMENTS

RESULT 1	
ID HV35_MOUSE	STANDARD: PRT: 111 AA.
AC P01804:	
DT 21-JUL-1986 (Rel. 01, Created)	
DT 21-JUL-1986 (Rel. 01, Last sequence update)	
DT 15-JUL-1999 (Rel. 38, Last annotation update)	
DE IG HEAVY CHAIN V-III REGION HPC76 (FRAGMENT).	
OS Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
RP MEDLINE=81013937; PubMed=6251474;	
RA Bernard O., Gough N.M.;	
RT "Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated VH and mu constant regions genes.";	
RL Proc. Natl. Acad. Sci. U.S.A. 77:3630-3634(1980).	
CC -I- MISCELLANEOUS: THE SEQUENCE OF THE FIRST 197 RESIDUES OF THE C REGION WAS ALSO DETERMINED & DIFFERS IN ONLY 3 POSITIONS FROM THE CORRESPONDING PORTION OF THE MOUSE MOPC 104E MU CHAIN.	
CC PIR: A02074; MMS76.	
DR InterPro: IPR003006; -	
DR Pfam: PF00047; 19; 1.	
KW Immunoglobulin V region.	
FT NON_TER 1	
FT NON_TER 111	
SQ SEQUENCE 111 AA; 12304 MW; 0EDE98ECT348056A CRC64;	
Query Match	83.7%; Score 36; DB 1; Length 111;
Best Local Similarity	87.5%; Pred. No. 2.1;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
OY 1 GRGLVPG 8	
DB 3 GGGLVPG 10	
RESULT 2	
ID HV27_MOUSE	STANDARD: PRT: 113 AA.
AC P01796:	
DT 21-JUL-1986 (Rel. 01, Created)	
DT 21-JUL-1986 (Rel. 01, Last sequence update)	
DT 15-JUL-1999 (Rel. 38, Last annotation update)	
DE IG HEAVY CHAIN V-III REGION A4.	
OS Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
RP MEDLINE=78158406; PubMed=417344;	
RX Vrana M., Rudikoff S., Potter M.;	

```

RT "Sequence variation among heavy chains from inulin-binding myeloma
RT proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS INULIN.
DR PIR: A93818; AVMSAB.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
KW Immunoglobulin V region.
FT DISULFID 22 98
FT NON_TER 113 113
FT SEQUENCE 113 AA; 12675 MW; 76658C121C598285 CRC64;

OY 1 GGGVOPG 8
Db 8 GGGVOPG 15

Query Match 83.7%; Score 36; DB 1; Length 113;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 3
HV28_MOUSE STANDARD; PRT; 113 AA.
AC P01797;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-III REGION U61.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=78158406; PubMed=417344;
RA Vrana M., Rudikoff S., Potter M.;
RT "Sequence variation among heavy chains from inulin-binding myeloma
RT proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS INULIN.
DR PIR: B93818; AVMS61.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
KW Immunoglobulin V region.
FT DISULFID 22 98
FT NON_TER 113 113
FT SEQUENCE 113 AA; 12671 MW; 76634C121C598285 CRC64;

OY 1 GGGVOPG 8
Db 8 GGGVOPG 15

Query Match 83.7%; Score 36; DB 1; Length 113;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 4
HV29_MOUSE STANDARD; PRT; 113 AA.
AC P01798;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-III REGION E109.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=78158406; PubMed=417344;
RA Vrana M., Rudikoff S., Potter M.;
RT "Sequence variation among heavy chains from inulin-binding myeloma
RT proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 75:1957-1961(1978).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS INULIN.
DR PIR: C93818; AVMS09.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
KW Immunoglobulin V region.
FT DISULFID 22 98
FT NON_TER 113 113
FT SEQUENCE 113 AA; 12647 MW; EE50F2F20EDB129B CRC64;

OY 1 GGGVOPG 8
Db 8 GGGVOPG 15

Query Match 83.7%; Score 36; DB 1; Length 113;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 5
HV30_MOUSE STANDARD; PRT; 113 AA.
AC P01799;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-III REGION ABE-47N.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=77134726; PubMed=402936;
RA Vrana M., Rudikoff S., Potter M.;
RT "Heavy-chain variable-region sequence from an inulin-binding myeloma
RT protein."
RL Biochemistry 16:1170-1175(1977).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS INULIN.
DR PIR: A90400; AVMS57.
DR InterPro: IPR003006; -.
DR Pfam: PF00047; 1g; 1.
KW Immunoglobulin V region.
FT DISULFID 22 98
FT NON_TER 113 113
FT SEQUENCE 113 AA; 12675 MW; 76658C16C779845E CRC64;

OY 1 GGGVOPG 8
Db 8 GGGVOPG 15

Query Match 83.7%; Score 36; DB 1; Length 113;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 6
HV31_MOUSE STANDARD; PRT; 113 AA.
AC P01800;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)

```

DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION T957.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE=81216632; PubMed=6787122;  
 RA Rudnikoff S., Potter M.;  
 RT "Immunoglobulin heavy chains from anti-inulin myeloma proteins:  
 RT evidence for a new heavy chain joining segment.";  
 RL J. Immunol. 127:191-194(1981).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: A92810; AVMS57.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113 BY SIMILARITY.  
 SQ SEQUENCE 113 AA; 12732 MW; 26618f626b59859e CRC64;

Query Match 83.7%; Score 36; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 DB 8 GGGLVOPG 15

RESULT 7  
 HV34\_MOUSE  
 ID HV34\_MOUSE STANDARD; PRT; 113 AA.  
 AC P01803;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V REGION AMP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE=81216632; PubMed=6787122;  
 RA Rudnikoff S., Potter M.;  
 RT "Immunoglobulin heavy chains from anti-inulin myeloma proteins:  
 RT evidence for a new heavy chain joining segment.";  
 RL J. Immunol. 127:191-194(1981).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: A02073; HVMSM.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 113 113 BY SIMILARITY.  
 SQ SEQUENCE 113 AA; 12691 MW; 7a6d906aa966e9e CRC64;

Query Match 83.7%; Score 36; DB 1; Length 113;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 DB 8 GGGLVOPG 15

RESULT 8

HV32\_MOUSE  
 ID HV32\_MOUSE STANDARD; PRT; 115 AA.  
 AC P01801;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION J606.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE=82099361; PubMed=6798111;  
 RA Johnson N., Slankard J., Paul L., Hood L.;  
 RT "The complete V domain amino acid sequences of two myeloma Inulin-  
 RT binding proteins.";  
 RL J. Immunol. 128:302-307(1982).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: G92811; AVMS06.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 115 115 BY SIMILARITY.  
 SQ SEQUENCE 115 AA; 12810 MW; b67ad6638a121a5f CRC64;

Query Match 83.7%; Score 36; DB 1; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 DB 8 GGGLVOPG 15

RESULT 9  
 HV33\_MOUSE  
 ID HV33\_MOUSE STANDARD; PRT; 115 AA.  
 AC P01802;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE IG HEAVY CHAIN V-III REGION W3082.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE=82099361; PubMed=6798111;  
 RA Johnson N., Slankard J., Paul L., Hood L.;  
 RT "The complete V domain amino acid sequences of two myeloma Inulin-  
 RT binding proteins.";  
 RL J. Immunol. 128:302-307(1982).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC PIR: D92811; AVMS82.  
 DR InterPro: IPR003006; -.  
 DR Pfam: PF00047; 19; 1.  
 KW Immunoglobulin V region.  
 FT DISULFID 22 98  
 FT NON\_TER 115 115 BY SIMILARITY.  
 SQ SEQUENCE 115 AA; 12887 MW; 9b4517648c121c5a CRC64;

Query Match 83.7%; Score 36; DB 1; Length 115;  
 Best Local Similarity 87.5%; Pred. No. 2.2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8

Db 8 GGGVOPG 15

RESULT 10  
ID HV3D\_HUMAN STANDARD; PRT; 115 AA.

AC P01765;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE IG HEAVY CHAIN V-III REGION TIL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP MEDLINE=78005528; PubMed=409716;

RA Wang A.-C.; Wang I.Y.; Fudenberg H.H.;

RT "Immunoglobulin structure and genetics. Identity between variable

regions of a mu and a gamma2 chain.";

RL J. Biol. Chem. 252:7193-7199(1977).

CC -I- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS

OF IGH AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL

GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO

CC IDENTICAL.

PIR: A02048; H3HUTL.

DR InterPro: IPR003006; -

KM Pfam: PF00047; 1g. 1.

FT Immunoglobulin V region.

NON\_TER 115

SO SEQUENCE 115 AA; 12356 MW; 4DCC67D179F62326 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 115;

Best Local Similarity 87.5%; Pred. No. 2.2;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVOPG 8

Db 8 GGGVOPG 15

RESULT 11  
ID HV3E\_MOUSE STANDARD; PRT; 116 AA.

AC P01806;  
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE IG HEAVY CHAIN V REGION 441 PRECURSOR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP MEDLINE=82059449; PubMed=6795591;

RA Ollio R.; Auffray C.; Sikorav J.-L.; Rougeon F.;

RT "Mouse heavy chain variable regions: nucleotide sequence of a

germ-line Vh gene segment.";

RL Nucleic Acids Res. 9:4099-4109(1981).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; V00774; CAA24152.1; -

DR PIR: A02076; HVMS44.

DR HSSP: P01810; 2RFJ

DR InterPro: IPR003006; -

DR Pfam: PF00047; 1g. 1.

DE Immunoglobulin V region; Signal.

FT SIGNAL 1

FT CHAIN 19 116 IG HEAVY CHAIN V REGION 441.

FT NON\_TER 116

SO SEQUENCE 116 AA; 12911 MW; 6B4FF3EF9A95D9F3 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 116;

Best Local Similarity 87.5%; Pred. No. 2.2;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVOPG 8

Db 26 GGGVOPG 33

RESULT 12

ID HV3R\_HUMAN STANDARD; PRT; 116 AA.

AC P01779;  
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE IG HEAVY CHAIN V-III REGION TUR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP MEDLINE=74142702; PubMed=4522293;

RA Capra J.D.; Kehoe J.M.;

RT "Variable region sequences of five human immunoglobulin heavy chains

RT of the Vh3 subgroup: definitive identification of four heavy chain

RT hypervariable regions.";

RL Proc. Natl. Acad. Sci. U.S.A. 71:845-848(1974).

CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGAL MYELOMA

CC PROTEIN.

PIR: A02062; A1HUTV.

DR InterPro: IPR003006; -

KM Pfam: PF00047; 1g. 1.

FT Immunoglobulin V region.

NON\_TER 116

SO SEQUENCE 116 AA; 12431 MW; EB705F553A963F0C CRC64;

Query Match 83.7%; Score 36; DB 1; Length 116;

Best Local Similarity 87.5%; Pred. No. 2.2;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVOPG 8

Db 8 GGGVOPG 15

RESULT 13

ID HV17\_MOUSE STANDARD; PRT; 117 AA.

AC P01786;  
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE IG HEAVY CHAIN V REGION WOPC 47A.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP MEDLINE=74142702; PubMed=4522293;

RX MEDLINE=80049769; PubMed=115869;  
RA Robinson E.A.; Appella E.;  
RT "Antino acid sequence of a mouse myeloma immunoglobulin heavy chain  
RL (MOPC 47 A) with a 100-residue deletion.";  
J. Biol. Chem. 254:11418-11430(1979).  
CC -1- MISCELLANEOUS: THIS ALPHA CHAIN WAS ISOLATED FROM A MYELOMA  
PROTEIN THAT CONTAINS ONE LIGHT AND ONE HEAVY CHAIN PER MOLECULE,  
CC LINKED BY A DISULFIDE BOND. IN CONTRAST, NORMAL MOUSE IGA  
CC MOLECULES CONTAIN TWO LIGHT AND TWO HEAVY CHAINS AND LACK A  
CC LIGHT-HEAVY CHAIN DISULFIDE BOND.  
DR PIR; A02069; AIMS47.  
DR HSSP; P01789; 2MCP.  
DR InterPro: IPR003006; -  
DR Pfam; PF00047; 1g: 1.  
KW Immunoglobulin V region.  
FT NON\_TER 117  
SQ SEQUENCE 117 AA; 12975 MW; 0C74BE8B154BDF4 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 117;  
Best Local Similarity 87.5%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
1 | 111111  
DB 8 GGGLVOPG 15

RESULT 14  
HV3C\_HUMAN STANDARD; PRT; 117 AA.  
ID P01764;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION VH26 PRECURSOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81101090; PubMed=6450418;  
RA Matthysens G.; Rabbitts T.H.;  
RT "Structure and multiplicity of genes for the human immunoglobulin  
RL heavy chain variable region.";  
Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J00236; AAAS3516.1; -  
DR EMBL; M33415; AAAS8735.1; -  
DR PIR; A02047; H3H026.  
DR InterPro: IPR003006; -  
DR Pfam; PF00047; 1g: 1.  
KW Immunoglobulin V region; Signal.  
FT SIGNAL 1  
FT CHAIN 20 117 IG HEAVY CHAIN V-III REGION VH26.  
FT NON\_TER 117  
SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB0F1 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 117;  
Best Local Similarity 87.5%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8

DB 1 | 111111  
27 GGGLVOPG 34

RESULT 15  
HV3C\_HUMAN STANDARD; PRT; 117 AA.  
ID P01764;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IG HEAVY CHAIN V-III REGION WAS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=74142702; PubMed=4522793;  
RA Capra J.D.; Kenne J.M.;  
RT "Variable region sequences of five human immunoglobulin heavy chains  
RT of the VH3 subgroup: definitive identification of four heavy chain  
RT hypervariable regions.";  
Proc. Natl. Acad. Sci. U.S.A. 71:845-848(1974).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA  
CC PROTEIN.  
DR PIR; A02059; G1H0WS.  
DR HSSP; P01772; 2IG2.  
DR InterPro: IPR003006; -  
DR Pfam; PF00047; 1g: 1.  
KW Immunoglobulin V region.  
FT NON\_TER 117  
SQ SEQUENCE 117 AA; 13091 MW; 201DEF0E1E53D9BF CRC64;

Query Match 83.7%; Score 36; DB 1; Length 117;  
Best Local Similarity 87.5%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
1 | 111111  
DB 8 GGGLVOPG 15

Search completed: June 13, 2001, 14:21:44  
Job time: 804 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:20:30 ; Search time 225.85 Seconds  
(without alignments)  
4.152 Million cell updates/sec

Title: PCT-US01-05825A-11  
Perfect score: 43  
Sequence: 1 GRGLVQPG 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_15:  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.unclassified:\*  
13: sp.vertebrate:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37	86.0	197	5	09N3C9	09n3c9 caenorhabd
2	36	83.7	15	11	09QV16	09qv16 rattus sp.
3	36	83.7	95	4	09ULB6	09ulb6 homo sapien
4	36	83.7	118	4	09UL91	09ul91 homo sapien
5	36	83.7	118	4	09UL72	09ul72 homo sapien
6	36	83.7	736	14	056137	056137 adeno-assoc
7	36	83.7	736	14	09WBP8	09wbp8 adeno-assoc
8	36	83.7	1439	13	09YGS4	09ygs4 brachydanto
9	35	81.4	16	4	09UC53	09uc53 homo sapien
10	35	81.4	104	11	09JLMI	09jlm1 mus musculu
11	35	81.4	198	4	09NPS2	09nps2 homo sapien
12	35	81.4	398	10	080872	080872 arabidopsis
13	35	81.4	514	11	09R266	09r266 mus musculu
14	35	81.4	1128	11	09QXZ4	09qxz4 mesocricetu
15	34	79.1	295	2	09K9C9	09k9c9 bacillus ha
16	34	79.1	404	5	022030	022030 caenorhabd1
17	34	79.1	433	4	09RY94	09ry94 delinococcus
18	34	79.1	457	4	043189	043189 homo sapien
19	34	79.1	713	4	075325	075325 homo sapien

20	34	79.1	714	2	051753	051753 pseudomonas
21	34	79.1	847	2	09RRN5	09rrn5 delinococcus
22	33	76.7	102	10	023369	023369 arabidopsis
23	33	76.7	113	4	09UL90	09ul90 homo sapien
24	33	76.7	116	4	09UL93	09ul93 homo sapien
25	33	76.7	121	4	09UL71	09ul71 homo sapien
26	33	76.7	122	4	09UL84	09ul84 homo sapien
27	33	76.7	147	4	09Y509	09y509 homo sapien
28	33	76.7	168	2	09S2V7	09s2v7 streptomyce
29	33	76.7	306	4	099848	099848 homo sapien
30	33	76.7	317	2	052286	052286 agrobacteri
31	33	76.7	325	11	09JHE4	09jhe4 mus musculu
32	33	76.7	329	4	09S344	09s344 homo sapien
33	33	76.7	367	2	032955	032955 mycobacteri
34	33	76.7	367	5	09V427	09v427 drosophila
35	33	76.7	421	4	09NP21	09np21 homo sapien
36	33	76.7	423	11	09JHE4	09jhe4 mus musculu
37	33	76.7	445	4	09NP18	09np18 homo sapien
38	33	76.7	446	4	09NP19	09np19 homo sapien
39	33	76.7	451	4	09NP20	09np20 homo sapien
40	33	76.7	455	4	09UP88	09up88 homo sapien
41	33	76.7	456	4	09Y215	09y215 homo sapien
42	33	76.7	542	2	09S3V2	09s3v2 chromobacte
43	33	76.7	552	5	09VFB5	09vfb5 drosophila
44	33	76.7	581	2	051863	051863 escherichia
45	33	76.7	581	2	09S6H9	09s6h9 escherichia

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	197 AA.
09N3C9	09N3C9	PRELIMINARY:	PRT:	197 AA.
AC	09N3C9;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	Y54E10BR.C PROTEIN.			
GN	Y54E10BR.C.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE=99069613; PubMed=9851916;			
RA	None;			
RT	"Genome sequence of the nematode C. elegans: a platform for			
RT	investigating biology. The C. elegans Sequencing Consortium."			
RL	Science 282:2012-2018(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Waterston R.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AC024812; AAF59550.1; -			
SQ	SEQUENCE 197 AA; 22094 MW; 97355DD1632AF872 CRC64;			
Query Match	Best Local Similarity	86.0%;	Score 37;	DB 5; Length 197;
Matches	6; Conservative	1;	Mismatches	1; Indels 0; Gaps 0;
OY	1 GRGLVQPG 8			
DB	78 GRGLVQPG 85			
RESULT	2	PRELIMINARY:	PRT:	15 AA.
09QV16	09QV16	PRELIMINARY:	PRT:	15 AA.
ID	09QV16	PRELIMINARY:	PRT:	15 AA.

AC Q90V16;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
 DE PROLACTIN-BINDING PROTEIN (FRAGMENT).  
 OS Rattus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10118;  
 RN [1]  
 RP MEDLINE=95094032; PubMed=8000909;  
 RX Cohen H., Cohen O., Gagnon J.:  
 RT "Serum prolactin-binding protein (PRL-BP) of human and rat are  
 RT identified as IgG.";  
 RL C. R. Acad. Sci., III, Sci. Vie 317:293-298(1994).  
 DR HSSP: P01789; IMCP.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 15 AA; 1469 MW; 35ED2512FF3FA369 CRC64;

Query Match 83.7%; Score 36; DB 11; Length 15;  
 Best Local Similarity 87.5%; Pred. NO. 1.2;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVQPG 8  
 Db 8 GGGVQPG 15

RESULT 3  
 ID Q9ULB6 PRELIMINARY; PRT; 95 AA.  
 AC Q9ULB6;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
 DE IMMUNOGLOBULIN HEAVY CHAIN (FRAGMENT).  
 GN VH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tange Y., Kayano H.:  
 RT "Human VH gene sequence.";  
 RL Submitted (NCV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB035268; BAA87067.1; -.  
 DR HSSP: P01772; 2FB4.  
 DR INTERPRO: IPR003006; -.  
 DR PFAM: PF00047; 19; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 95 AA; 10527 MW; 90A8C6D16D2574A CRC64;

Query Match 83.7%; Score 36; DB 4; Length 95;  
 Best Local Similarity 87.5%; Pred. NO. 8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVQPG 8  
 Db 7 GGGVQPG 14

RESULT 4  
 ID Q9UL91 PRELIMINARY; PRT; 118 AA.  
 AC Q9UL91;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, last annotation update)  
 DE MOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.:  
 RT "Mosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035023; AAD56259.1; -.  
 DR INTERPRO: IPR003006; -.  
 DR PFAM: PF00047; 19; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match 83.7%; Score 36; DB 4; Length 118;  
 Best Local Similarity 87.5%; Pred. NO. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVQPG 8  
 Db 8 GGGVQPG 15

RESULT 5  
 ID Q9UL72 PRELIMINARY; PRT; 118 AA.  
 AC Q9UL72;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
 DE MOSIN-REACTIVE IMMUNOGLOBULIN HEAVY CHAIN VARIABLE REGION (FRAGMENT).  
 GN VH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.:  
 RT "Mosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035042; AAD56278.1; -.  
 DR HSSP: P01772; 2FB4.  
 DR INTERPRO: IPR003006; -.  
 DR PFAM: PF00047; 19; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 118 AA; 12872 MW; B4D1A5944B2D5CCA CRC64;

Query Match 83.7%; Score 36; DB 4; Length 118;  
 Best Local Similarity 87.5%; Pred. NO. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVQPG 8  
 Db 8 GGGVQPG 15

RESULT 6  
 ID O56137 PRELIMINARY; PRT; 736 AA.  
 AC O56137;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, last sequence update)



```

DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE CAPSID PROTEIN VP1.
OS adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68538;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
  (AAV) serotypes other than AAV type 2."
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1; -.
DR INTERPRO: IPR001403; -.
DR PFAM: PF00740; Parvo_coat.1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;

Query Match
Best Local Similarity 83.7%; Score 36; DB 14; Length 736;
Pred. NO. 68;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8
    |||||
Db 42 GRGLVPG 49

RESULT 7
ID Q9MBP8 PRELIMINARY; PRT; 736 AA.
AC Q9MBP8;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE CAPSID PROTEIN.
OS adeno-associated virus 1.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=85106;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214338; PubMed=10196295;
RA Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;
RT "Gene therapy vectors based on adeno-associated virus type 1."
RL J. Virol. 73:3994-4003(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Xiao W., Wilson J.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063497; AAD27757.1; -.
DR INTERPRO: IPR001403; -.
DR PFAM: PF00740; Parvo_coat.1.
SQ SEQUENCE 736 AA; 81375 MW; CFA8FB9BD5CD0595 CRC64;

Query Match
Best Local Similarity 83.7%; Score 36; DB 14; Length 736;
Pred. NO. 68;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8
    |||||
Db 42 GRGLVPG 49

RESULT 8
ID Q9YCS4 PRELIMINARY; PRT; 1439 AA.
AC Q9YCS4;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

```

```

DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE ZINC FINGER TRANSCRIPTION FACTOR GLI2.
CN YOI-TOO.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99158735; PubMed=10049354;
RA Karlstrom R.O., Talbot W.S., Schier A.F.;
RT "Comparative synteny cloning of zebrafish you-too: mutations in the
  Hedgehog target glis2 affect ventral forebrain patterning."
RL Genes Dev. 13:388-393(1999).
DR EMBL; AF085746; AAD18135.2; -.
DR HSSP; P08151; 2GLI.
DR INTERPRO: IPR000822; -.
DR INTERPRO: IPR002057; -.
DR PFAM: PF00096; zf-C2H2_5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_4.
DR PROSITE; PS00185; IPNS_1; 1.
KW Zinc-finger; Metal-binding; DNA-binding.
SQ SEQUENCE 1439 AA; 156790 MW; C52934C05FDF3A3FA CRC64;

Query Match
Best Local Similarity 83.7%; Score 36; DB 13; Length 1439;
Pred. NO. 1,46+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGLVOPG 7
    |||||
Db 1248 GRGLVPG 1254

RESULT 9
ID Q9UC53 PRELIMINARY; PRT; 16 AA.
AC Q9UC53;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 14, Last annotation update)
DE 77 KDA SPONTANEOUS RECURRENT ABORTION-ASSOCIATED HUMAN EMBRYONIC
DE ANTIGEN/IGWHIII HOMOLOG (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=96033130; PubMed=8582963;
RA Shirasahi Y., Shirasahi Y., Yamamoto D., Hasegawa T., Kitamura W.,
RA Miki S., Tanaka T., Suzuki T., Soma H.;
RT "Diagnostic relevance of abortion-associated human embryonic antigen
  expressed on the cell surface of tumour promoter-treated Bloom
  syndrome cells."
RL Hum. Reprod. 10:1694-1701(1995).
SQ SEQUENCE 16 AA; 1626 MW; C9C5ED2512FF3B9 CRC64;

Query Match
Best Local Similarity 81.4%; Score 35; DB 4; Length 16;
Pred. NO. 1.9;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8
    |||||
Db 8 GRGLVPG 15

RESULT 10
ID Q9J1M1 PRELIMINARY; PRT; 104 AA.
AC Q9J1M1;
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

```

DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE TELOMERASE REVERSE TRANSCRIPTASE (FRAGMENT).  
 GN TERT.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hatakeyama S., Ishikawa F.;  
 RT "Identification of the mouse telomerase reverse transcriptase (mTERT)  
 RT promoter.";  
 RT Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF151502; AAF42984.1; -;  
 KW RNA-directed DNA polymerase.  
 FT NON\_TER 104  
 FT 104  
 SQ SEQUENCE 104 AA; 11998 MW; 4B649E63476D3D44 CRC64;

Query Match 81.4%; Score 35; DB 11; Length 104;  
 Best Local Similarity 87.5%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 |||||  
 DB 35 GRGLVOPG 42

RESULT 11  
 Q9NPS2 PRELIMINARY; PRT; 198 AA.  
 AC Q9NPS2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 21.4 KDA PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Carim L., Estivill X., Escarceller M., Sumoy L.;  
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Aufray C., Ansoy W., Ballabio A., Estivill X., Gibson K.,  
 RA Leheach H., Poustka A., Lundberg J.;  
 RT "The European IMAGE consortium for Integrated Molecular analysis of  
 RT human gene transcripts.";  
 RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AL365410; CAB96954.1; -;  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT 1  
 SQ SEQUENCE 198 AA; 21446 MW; D54853C4A0A4CAF5 CRC64;

Query Match 81.4%; Score 35; DB 4; Length 198;  
 Best Local Similarity 75.0%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 |||||  
 DB 104 GRGLVOPG 111

RESULT 12  
 ID 080872 PRELIMINARY; PRT; 398 AA.  
 AC 080872;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)

DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE F23F1.7 PROTEIN.  
 GN F23F1.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,  
 RA Somerville C.R., Venter J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.";  
 RT Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC004680; AAC31851.1; -;  
 SQ SEQUENCE 398 AA; 45476 MW; F55EBBFA49C1C1D9 CRC64;

Query Match 81.4%; Score 35; DB 10; Length 398;  
 Best Local Similarity 62.5%; Pred. No. 56;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 |||||  
 DB 78 GRGLVOPG 85

RESULT 13  
 Q9R266 PRELIMINARY; PRT; 514 AA.  
 AC Q9R266;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE TELOMERASE REVERSE TRANSCRIPTASE (FRAGMENT).  
 GN TERT.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99144726; PubMed=10022128;  
 RA Greenberg R.A., O'Hagan R.C., Deng H., Xiao Q., Hann S.R., Adams R.R.,  
 RA Lichtsteiner S., Chin L., Morin G.B., DePinho R.A.;  
 RT "Telomerase reverse transcriptase gene is a direct target of c-Myc but  
 RT is not functionally equivalent in cellular transformation.";  
 RL Oncogene 18:1219-1226(1999).  
 DR EMBL; AF121949; AAD24465.1; -;  
 KW RNA-directed DNA polymerase.  
 FT NON\_TER 514  
 FT 514  
 SQ SEQUENCE 514 AA; 58382 MW; 3096599776D9BBD CRC64;

Query Match 81.4%; Score 35; DB 11; Length 514;  
 Best Local Similarity 87.5%; Pred. No. 73;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GRGLVOPG 8  
 |||||  
 DB 35 GRGLVOPG 42

RESULT 14  
 ID 090XZ4 PRELIMINARY; PRT; 1128 AA.  
 AC 090XZ4;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)

DE TELOMERASE CATALYTIC SUBUNIT.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Guo W., Okamoto M., Baluda M.A., Park N.-H.;  
 RT "Cloning of the cDNA and promoter of hamster telomerase catalytic  
 subunit (hamTERT)."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF149012; AAF17334.1;  
 SQ SEQUENCE 1128 AA; 128393 MW; 1D4F81249012174E CRC64;

Query Match 81.4%; Score 35; DB 11; Length 1128;  
 Best Local Similarity 87.5%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQPG 8  
 |||||  
 DB 35 GRGLVQPG 42

RESULT 15  
 ID 09K9C9 PRELIMINARY: PRT: 295 AA.  
 AC 09K9C9;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE BH2720. PROTEIN.  
 GN BH2720.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RA Takami H., Nakasone K., Takaki Y.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP001516; BAB06439.1;  
 SQ SEQUENCE 295 AA; 34308 MW; 1BDE317EDA39B764 CRC64;

Query Match 79.1%; Score 34; DB 2; Length 295;  
 Best Local Similarity 87.5%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQPG 8  
 |||||  
 DB 220 GRGLVQMG 227

Search completed: June 13, 2001, 14:20:31  
 Job time: 731 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 14:16:34 ; Search time 118.55 Seconds  
(without alignments)  
1.296 Million cell updates/sec

Title: PCT-US01-05825A-11

Perfect score: 43

Sequence: 1 GRGLVPG 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	246	1	US-07-843-125-11 Sequence 11, Appl
2	38	88.4	116	3	US-08-545-809A-143 Sequence 143, Appl
3	37	86.0	118	2	US-08-652-816A-112 Sequence 12, Appl
4	36	83.7	15	1	US-08-331-398A-66 Sequence 66, Appl
5	36	83.7	15	2	US-08-652-558-42 Sequence 42, Appl
6	36	83.7	15	2	US-08-331-397B-66 Sequence 66, Appl
7	36	83.7	15	2	US-08-759-804A-65 Sequence 65, Appl
8	36	83.7	18	1	US-08-331-398A-55 Sequence 55, Appl
9	36	83.7	18	1	US-08-401-908-14 Sequence 14, Appl
10	36	83.7	18	2	US-08-331-397B-55 Sequence 55, Appl
11	36	83.7	18	2	US-08-759-804A-54 Sequence 54, Appl
12	36	83.7	20	2	US-08-859-931A-2 Sequence 2, Appl
13	36	83.7	20	2	US-08-053-451B-113 Sequence 113, Appl
14	36	83.7	26	1	US-08-471-780C-80 Sequence 80, Appl
15	36	83.7	26	1	US-08-467-282B-80 Sequence 80, Appl
16	36	83.7	26	2	US-08-471-282A-80 Sequence 80, Appl
17	36	83.7	26	2	US-08-466-710C-80 Sequence 80, Appl
18	36	83.7	26	3	US-08-468-739C-80 Sequence 80, Appl
19	36	83.7	29	1	US-08-471-780C-119 Sequence 119, Appl
20	36	83.7	29	1	US-08-467-282B-119 Sequence 119, Appl
21	36	83.7	29	2	US-08-471-282A-119 Sequence 119, Appl
22	36	83.7	29	2	US-08-466-710C-119 Sequence 119, Appl
23	36	83.7	29	3	US-08-468-739C-119 Sequence 119, Appl
24	36	83.7	30	1	US-07-988-925-7 Sequence 7, Appl
25	36	83.7	30	1	US-07-977-696C-75 Sequence 75, Appl
26	36	83.7	30	1	US-08-129-930B-75 Sequence 75, Appl
27	36	83.7	30	2	US-08-362-780-7 Sequence 7, Appl

## ALIGNMENTS

28	36	83.7	30	2	US-08-765-783A-92	Sequence 92, Appl
29	36	83.7	30	2	US-08-470-139-17	Sequence 17, Appl
30	36	83.7	67	1	US-08-162-102C-36	Sequence 36, Appl
31	36	83.7	86	2	US-08-053-451B-126	Sequence 126, Appl
32	36	83.7	87	1	US-08-497-312-21	Sequence 21, Appl
33	36	83.7	87	2	US-08-765-783A-105	Sequence 105, Appl
34	36	83.7	98	1	US-08-211-202-116	Sequence 116, Appl
35	36	83.7	98	1	US-07-942-245-37	Sequence 37, Appl
36	36	83.7	98	2	US-08-428-197-48	Sequence 48, Appl
37	36	83.7	98	2	US-08-665-202-31	Sequence 31, Appl
38	36	83.7	98	2	PCT-US93-10555-48	Sequence 48, Appl
39	36	83.7	108	2	US-08-428-197-4	Sequence 4, Appl
40	36	83.7	108	5	PCT-US93-10555-4	Sequence 4, Appl
41	36	83.7	109	1	US-07-942-245-17	Sequence 17, Appl
42	36	83.7	109	1	US-07-942-245-21	Sequence 21, Appl
43	36	83.7	109	2	US-08-428-197-3	Sequence 3, Appl
44	36	83.7	109	5	PCT-US93-10555-3	Sequence 3, Appl
45	36	83.7	110	1	US-08-211-202-117	Sequence 117, Appl

RESULT 1  
US-07-843-125-11  
; Sequence 11, Application US/07843125  
; Patent No. 5395750  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J  
; TITLE OF INVENTION: Methods for Producing Proteins which  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; City: Nutley  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; Zip: 07110-1199  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/843,125  
; FILING DATE: 19920228  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roseman, Catherine R  
; REGISTRATION NUMBER: 34240  
; REFERENCE/DOCKET NUMBER: 8594  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-6208  
; TELEFAX: (201) 235-3500  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 246 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-843-125-11

Query Match 100.0%; Score 43; DB 1; Length 246;  
Best Local Similarity 100.0%; Pred. No. 0.76;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGLVPG 8  
Db 8 GRGLVPG 15

RESULT 2  
US-08-545-809A-143  
Sequence 143, Application US/08545809A  
Patent No. 6096878  
GENERAL INFORMATION:  
APPLICANT: Honjo, Yasuko  
APPLICANT: Matsuda, Fumihiko  
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE  
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME  
NUMBER OF SEQUENCES: 145  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: PASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545, 809A  
FILING DATE: 27-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/00603  
FILING DATE: 10-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 06501/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-809A-143

Query Match 88.4%; Score 38; DB 3; Length 116;  
Best Local Similarity 87.5%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVQPG 8  
DB 27 GRLVQPG 34

RESULT 3  
US-08-652-816A-12  
Sequence 12, Application US/08652816A  
Patent No. 5872215  
GENERAL INFORMATION:  
APPLICANT: Osbourn, JK  
APPLICANT: Allen, DJ  
APPLICANT: McCallister, JG  
TITLE OF INVENTION: Specific binding members, materials and  
TITLE OF INVENTION: methods.  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652, 816A  
FILING DATE: 23-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9125579.4  
FILING DATE: 02-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9125579.8  
FILING DATE: 02-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206372.6  
FILING DATE: 23-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9525004.9  
FILING DATE: 07-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9610824.6  
FILING DATE: 23-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/02240  
FILING DATE: 02-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/244,597  
FILING DATE: 01-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: David W. Clough  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/33308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-652-816A-12

Query Match 86.0%; Score 37; DB 2; Length 118;  
Best Local Similarity 87.5%; Pred. No. 4.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRLVQPG 8  
DB 8 GRLVQPG 15

RESULT 4  
US-08-331-398A-66  
Sequence 66, Application US/08331398A  
Patent No. 5608039  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David  
APPLICANT: Brinkmann, Ulrich  
APPLICANT: Pal, Lee  
TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins  
TITLE OF INVENTION: and their uses (as amended)  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Steuart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,398A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-398A-66

Query Match 83.7% Score 36; DB 1; Length 15;  
Best Local Similarity 87.5% Pred. No. 0.86;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGVOPG 8  
| | | | |  
DB 8 GGGVOPG 15

RESULT 5  
US-08-652-558-42  
Sequence 42, Application US/08652558  
Patent No. 5861155  
GENERAL INFORMATION:  
APPLICANT: LIN, AUGUSTINE YEE-THARN  
TITLE OF INVENTION: HUMANIZED ANTIBODIES AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & WITCOFF  
STREET: 75 STATE STREET, 23RD FLOOR  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,558  
FILING DATE: JUNE 6, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/IB94/00387  
FILING DATE: NOVEMBER 21, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YANKWICH, LEON R.  
REGISTRATION NUMBER: 30,237  
REFERENCE/DOCKET NUMBER: 95,497-L

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-345-9100  
TELEFAX: 617-345-9111  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-558-42

Query Match 83.7% Score 36; DB 2; Length 15;  
Best Local Similarity 87.5% Pred. No. 0.86;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGVOPG 8  
| | | | |  
DB 8 GGGVOPG 15

RESULT 6  
US-08-331-397B-66  
Sequence 66, Application US/08331397B  
Patent No. 5981726  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Benhar, Ital  
TITLE OF INVENTION: Chimeric and Mutationally Stabilized Tumor-Specific Antibody Fragments, Fusion Proteins, and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,397B  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126120US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-397B-66

Query Match 83.7% Score 36; DB 2; Length 15;

Best Local Similarity 87.5%; Pred. No. 0.86;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGLVOPG 8  
Db 8 GCGLVOPG 15

RESULT 7  
US-08-759-804A-65  
; Sequence 65, Application US/08759804A  
; Patent No. 5990296  
; GENERAL INFORMATION:  
; APPLICANT: Pastan, Ira  
; APPLICANT: Willingham, Mark  
; APPLICANT: Fitzgerald, David J.  
; APPLICANT: Brinkmann, Ulrich  
; APPLICANT: Pal, Lee  
; TITLE OF INVENTION: Tumor-Specific Antibody Fragments,  
; TITLE OF INVENTION: Fusion Proteins, and Uses Thereof  
; NUMBER OF SEQUENCES: 68  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/759,804A  
; FILING DATE: 03-DEC-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/331,398  
; FILING DATE: 28-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/767,331  
; FILING DATE: 30-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/596,289  
; FILING DATE: 12-OCT-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen L.  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 015280-1261400S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-759-804A-65

Query Match 83.7%; Score 36; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 0.86;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGLVOPG 8  
Db 8 GCGLVOPG 15

RESULT 8

US-08-331-398A-55  
; Sequence 55, Application US/08331398A  
; Patent No. 5608039

; GENERAL INFORMATION:  
; APPLICANT: Pastan, Ira  
; APPLICANT: Willingham, Mark  
; APPLICANT: Fitzgerald, David  
; APPLICANT: Brinkmann, Ulrich  
; APPLICANT: Pal, Lee  
; TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins  
; TITLE OF INVENTION: and Their Uses (as amended)  
; NUMBER OF SEQUENCES: 68  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: One Market Plaza, Stewart Street Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/331,398A  
; FILING DATE: 28-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/767,331  
; FILING DATE: 30-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/596,289  
; FILING DATE: 12-OCT-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 015280-1261100S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 55:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-331-398A-55

Query Match 83.7%; Score 36; DB 1; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGLVOPG 8  
Db 8 GCGLVOPG 15

RESULT 9  
US-08-401-908-14  
; Sequence 14, Application US/08401908  
; Patent No. 5684146  
; GENERAL INFORMATION:  
; APPLICANT: Yoshinobu OKUNO et al.  
; TITLE OF INVENTION: A DNA WHICH CODES FOR THE VARIABLE REGION OF  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington



STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,908  
FILING DATE: March 10, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE: heavy chain of C179  
US-08-401-908-14

Query Match 83.7%; Score 36; DB 1; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVOPG 8  
DB 8 GGGVOPG 15

RESULT 10  
US-08-331-397B-55  
Sequence 55, Application US/08331397B  
Patent No. 5981726  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Benhar, Itai  
TITLE OF INVENTION: Chimeric and Mutationally Stabilized Tumor-Specific Antibody Fragments, Fusion Proteins, and Uses  
TITLE OF INVENTION: Theoreof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,397B  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-12612005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-397B-55

Query Match 83.7%; Score 36; DB 2; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGVOPG 8  
DB 8 GGGVOPG 15

RESULT 11  
US-08-759-804A-54  
Sequence 54, Application US/08759804A  
Patent No. 5990296  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Willingham, Mark  
APPLICANT: Fitzgerald, David J.  
APPLICANT: Brinkmann, Ulrich  
TITLE OF INVENTION: Tumor-Specific Antibody Fragments, Fusion Proteins, and Uses Thereof  
TITLE OF INVENTION: Theoreof  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/759,804A  
FILING DATE: 03-DEC-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/331,398  
FILING DATE: 28-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 015280-12614005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 54:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-759-804A-54

Query Match 83.7%; Score 36; DB 2; Length 18;  
Best Local Similarity 87.5%; Pred. No. 1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGVOPG 8  
| | | | |  
Db 8 GGGVOPG 15

RESULT 12  
US-08-859-931A-2  
; Sequence 2, Application US/08859931A  
; Patent No. 5945510  
; GENERAL INFORMATION:  
; APPLICANT: PASANO, Alessio  
; TITLE OF INVENTION: SUBSTANTIALLY PURE ZONULIN, A  
; TITLE OF INVENTION: PHYSIOLOGICAL MODULATOR OF  
; TITLE OF INVENTION: MAMMALIAN TIGHT JUNCTIONS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W., Suite 800  
; CITY: Washington, D.C.  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/859,931A  
; FILING DATE: 21 May 1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KIT, Gordon  
; REGISTRATION NUMBER: 30,764  
; REFERENCE/DOCKET NUMBER: A-6901  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 293-7060  
; TELEFAX: (202) 293-7860  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
US-08-859-931A-2

Query Match 83.7%; Score 36; DB 2; Length 20;  
Best Local Similarity 87.5%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGVOPG 8  
| | | | |  
Db 8 GGGVOPG 15

RESULT 13

US-08-053-451B-113  
; Sequence 113, Application US/08053451B  
; Patent No. 5955584  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Francis W.  
; APPLICANT: Dittow, Charles C.  
; APPLICANT: Calenoff, Emanuel  
; TITLE OF INVENTION: ATHEROSCLEROTIC PLAQUE SPECIFIC  
; TITLE OF INVENTION: ANTIGENS, ANTIBODIES THEREOF, AND USES THEREOF  
; NUMBER OF SEQUENCES: 176  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/053,451B  
; FILING DATE: 26-APR-1993  
; CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
; NAME: Halluin, Albert P.  
; REGISTRATION NUMBER: 25,227  
; REFERENCE/DOCKET NUMBER: 7606-033-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-854-3694  
; TELEFAX: 415-854-3694  
; TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 113:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 2..3  
; OTHER INFORMATION: /note="Where Xaa = k or q = Lys or  
; OTHER INFORMATION: Glu"

US-08-053-451B-113

Query Match 83.7%; Score 36; DB 2; Length 20;  
Best Local Similarity 87.5%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGVOPG 8  
| | | | |  
Db 7 GGGVOPG 14

RESULT 14  
US-08-471-780C-80  
; Sequence 80, Application US/08471780C  
; Patent No. 5759808  
; GENERAL INFORMATION:  
; APPLICANT: Casterman, Cecile  
; APPLICANT: Hamers, Raymond  
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA

Query Match 83.7%; Score 36; DB 2; Length 20;  
Best Local Similarity 87.5%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,780C  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/106,944  
FILING DATE: 17-AUG-1993  
APPLICATION NUMBER: FR 92402326.0  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 93401310.3  
FILING DATE: 21-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E.R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 04958.0008-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Camelus dromedarius  
US-08-471-780C-80

Query Match 83.7%; Score 36; DB 1; Length 26;  
Best Local Similarity 87.5%; Pred. NO. 1.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQPG 8  
| | | | | | |  
DB 8 GGLVQPG 15

RESULT 15  
US-08-467-282B-80  
Sequence 80, Application US/08467282B  
Patent No. 5800988  
GENERAL INFORMATION:  
APPLICANT: Casleiman, Cecile  
APPLICANT: Hamers, Raymond  
TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,282B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/106,944

FILING DATE: 17-AUG-1993  
APPLICATION NUMBER: FR 92402326.0  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 93401310.3  
FILING DATE: 21-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E.R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 04958.0008-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Camelus dromedarius  
US-08-467-282B-80

Query Match 83.7%; Score 36; DB 1; Length 26;  
Best Local Similarity 87.5%; Pred. NO. 1.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRGLVQPG 8  
| | | | | | |  
DB 8 GGLVQPG 15

Search completed: June 13, 2001, 14:16:34  
Job time: 495 sec

